

STIC-Biotech/ChemLib

177 782

mej

From: Whiteman, Brian  
Sent: Friday, January 27, 2006 8:28 AM  
To: STIC-Biotech/ChemLib  
Subject: seq search

09/889,874

SEQ ID NO: 22 and 23

1) search us patents and published us patent application databases

Thank you,

Brian Whiteman  
Remsen, 2D14  
mail box 2C18  
Patent Examiner - Art Unit 1635  
United States Patent and Trademark Office  
(571) 272-0764

STIC  
JAN 27 2006  
10:00 AM

\*\*\*\*\*  
Point of Contact:  
Alexandra Wacławiw  
Technical Info. Specialist  
211 6A02 Tel 308-441  
Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: 1-30  
Date completed: 6-1-30  
Searcher Prep Time: \_\_\_\_\_  
Online Time: 8

\*\*\*\*\*  
Type of Search  
NA# \_\_\_\_\_ AA# 2  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*  
Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

THIS PAGE LEFT BLANK

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: January 30, 2006, 09:40:03 : Search time 25.6283 seconds  
(without alignments)  
5726.190 Million cell updates/sec

Title: US-09-889-874A-22

Perfect score: 1784  
Sequence: 1 FTLRBDSMDMTGVSTFNV.....ILFHAGYNSDPYLSSTIVY 334

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A\_Geneseq\_21:\*

- 1: Geneseq219808:\*
- 2: Geneseq219809:\*
- 3: Geneseq220008:\*
- 4: Geneseq220018:\*
- 5: Geneseq220028:\*
- 6: Geneseq220038:\*
- 7: Geneseq220048:\*
- 8: Geneseq220058:\*
- 9: Geneseq220068:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1784	100.0	334	3 AAY95706	AAY95706 Cosmid cH
2	145.5	8.2	341	6 ABM67282	ABM67282 Phototrab
3	126	7.1	892	6 ABU42557	ABU42557 Protein e
4	126	7.1	892	8 AD084849	AD084849 S. epiderm
5	126	7.1	892	8 AD084849	AD084849 S. epiderm
6	126	7.1	930	5 ABP40469	ABP40469 Staphyloc
7	126	7.1	930	5 ABP40469	ABP40469 Staphyloc
8	126	7.1	991	3 AAY83171	AAY83171 Cell wall
9	126	7.1	991	3 AAY70120	AAY70120 Staph. ep
10	124	7.0	1092	2 AAY41602	AAY41602 Staphyloc
11	124	7.0	1092	2 AAY41602	AAY41602 Staphyloc
12	119	6.7	893	7 ABM79019	ABM79019 Staphyloc
13	119	6.7	893	7 ABM79019	ABM79019 Staphyloc
14	118.5	6.6	899	7 ABH68812	ABH68812 Enterococ
15	118	6.6	899	7 ABH68812	ABH68812 Enterococ
16	115.5	6.5	773	6 ABB91631	ABB91631 Microbial
17	115.5	6.5	855	5 ABB98573	ABB98573 Dextran s
18	115.5	6.5	1411	6 ADA09358	ADA09358 Haemophil
19	115.5	6.5	1411	6 ADA09358	ADA09358 Haemophil
20	115.5	6.5	2835	5 ABR898574	ABR898574 Dextran s
21	114	6.4	360	6 ABR55594	ABR55594 Amino aci
22	114	6.4	360	6 ABR55594	ABR55594 Amino aci
23	114	6.4	376	8 ADA60413	ADA60413 Streptoco
24	112.5	6.3	1436	8 ADA09354	ADA09354 Haemophil

25	112.5	6.3	1436	8 ADRA6771	ADRA6771 H. influe
26	111	6.2	369	6 ABU01334	ABU01334 S. pneumo
27	109	6.1	807	3 AAB18311	AAB18311 Plasmodiu
28	109	6.1	807	7 ABO23607	ABO23607 Plasmodiu
29	108.5	6.1	753	8 ADU07705	ADU07705 Amino aci
30	108	6.1	626	3 AAG30820	AAG30820 Arabidops
31	108	6.1	626	3 AAG30820	AAG30820 Arabidops
32	108	6.1	978	6 ABU47432	ABU47432 Protein e
33	107.5	6.0	324	2 AAY19983	AAY19983 B. burgdo
34	107.5	6.0	343	2 AAY19982	AAY19982 B. burgdo
35	107.5	6.0	753	8 ADX91694	ADX91694 Plant ful
36	107.5	6.0	1038	6 ABU15903	ABU15903 Protein e
37	107	6.0	1072	5 ABB54963	ABB54963 Lactococ
38	107	6.0	1114	2 AAR21999	AAR21999 M17 anti
39	107	6.0	1449	6 ABB37787	ABB37787 Protein e
40	107	6.0	1449	6 ABB37787	ABB37787 Protein e
41	106.5	6.0	719	8 ADX93594	ADX93594 Plant ful
42	106	5.9	258	2 AAY33726	AAY33726 Phototrab
43	106	5.9	1430	5 AAY98042	AAY98042 S. mutans
44	105.5	5.9	991	8 ADU25478	ADU25478 L. acidop
45	105	5.9	507	3 AAB18173	AAB18173 Plasmodiu

## ALIGNMENTS

RESULT 1	AAY95706	AAY95706 standard; protein; 334 AA.
ID	AAY95706	AAY95706 standard; protein; 334 AA.
XX	XX	XX
AC	AAY95706;	
XX	XX	XX
DT	25-OCT-2000	(first entry)
XX	XX	XX
DE	Cosmid CHRM5 encoded protein P13-1f.	
XX	XX	XX
KW	Cosmid CHRM5; nematocite; nematode; biological control agent;	
XX	transgenic plant; helminthiasis; P13-1f.	
OS	Xenorhabdus bovienii.	
XX	XX	XX
PN	W0200042855-A1.	
XX	XX	XX
PD	27-JUL-2000.	
XX	XX	XX
PF	24-JAN-2000; 2000MO-GB000219.	
XX	XX	XX
PR	22-JAN-1999; 99GB-00001499.	
XX	XX	XX
PA	(HORT-) HORTICULTURE RES INT.	
XX	XX	XX
PT	Morgan JAW, Jarrett P, Ellis D, Ousley MA;	
XX	XX	XX
PT	WPI; 2000-499157/44.	
XX	XX	XX
DR	N-PSDB; AAY50029.	
XX	XX	XX
PT	Novel composition used to control parasitic nematodes, especially in a	
XX	plants such as maize, cotton, soya, and rice, comprises a bacterium which	
XX	is a symbiont of an entomopathogenic nematode.	
PS	Example 6; Page 42; 74pp; English.	
XX	XX	XX
CC	The present sequence is that of protein P13-1f encoded by an open reading	
CC	frame identified in cosmid CHRM5 (see AAY50029). CHRM5 was obtained by	
CC	ligating Xenorhabdus bovienii strain 173 (NCIMB 40966) Sau3A-digested DNA	
CC	fragments into the BamHI site of the Stratagene cosmid vector Supercos1,	
CC	packaging into Escherichia coli XL Blue 1, and screening for nematocidal	
CC	activity against Caenorhabditis elegans. Analysis of the DNA indicated a	
CC	number of open reading frames for which the corresponding protein	
CC	sequences were determined (see AAY5685-195735). Nematodes can be	
CC	controlled through the use of bacteria associated symbiotically with an	
CC	entomopathogenic nematode. Such bacteria include Xenorhabdus and	
CC	Photorhabdus spp. such as X. bovienii strain 173. The symbiont bacteria,	

CC an engineered bacterium, or a nematocidal protein obtained from such  
CC bacteria, particularly P13-1f or P14-2f (see AAY95707) can be used to  
CC control helminthiasis in a human or domesticated animal or for the  
CC control of plant pathogen nematodes. Also claimed are vectors for  
CC expressing nematocidal proteins in host cells, and transgenic plants  
XX

SQ Sequence 334 AA;

Query Match 100.0%; Score 1784; DB 3; Length 334;

Best Local Similarity 100.0%; Pred. No. 5.7e-150; Mismatches 0; Gaps 0;

Matches 334; Conservative 0; Indels 0; Gaps 0;

QY 1 FTLRSDMSDWTGVTENVILETGLDNCNIYANGLMIGVINITPTDDEGNFVDIDVT 60  
DB 1 FTLRSDMSDWTGVTENVILETGLDNCNIYANGLMIGVINITPTDDEGNFVDIDVT 60  
QY 61 LNDNIKIVDYIDGSDIDGSDGMPYTGPNENYITPNSQSYLKSNSQITQIKRYVSCS 120  
DB 61 LNDNIKIVDYIDGSDIDGSDGMPYTGPNENYITPNSQSYLKSNSQITQIKRYVSCS 120  
QY 121 NTSRLRKSFSKAKYTTSGKVISITONSINSRVVINAIDATNTDDELRTKRETNQ 180  
DB 121 NTSRLRKSFSKAKYTTSGKVISITONSINSRVVINAIDATNTDDELRTKRETNQ 180  
QY 181 SYTSKSTNSLYVHTWTIPRSLKLOWRMEDYNNNGWTWASQCYKXGADGSGSESTWLA 240  
DB 181 SYTSKSTNSLYVHTWTIPRSLKLOWRMEDYNNNGWTWASQCYKXGADGSGSESTWLA 240  
QY 241 AGSIFPPGNTDGLWLDNDIALSGMAHKSYNVDGTGINSFTRIIGKGSWYVNIISGLDRG 300  
DB 241 AGSIFPPGNTDGLWLDNDIALSGMAHKSYNVDGTGINSFTRIIGKGSWYVNIISGLDRG 300  
QY 301 HAVIIIDQYGNKRIILPHAGYENSDPYLSSSIYV 334  
DB 301 HAVIIIDQYGNKRIILPHAGYENSDPYLSSSIYV 334

RESULT 2

ABM67282 ID ABM67282 standard; protein; 341 AA.

XX ABM67282;

DT 20-NOV-2003 (first entry)

XX Photorhabdus luminescens protein sequence #379.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
KW detection; food; gene expression; plant; animal; microorganism; toxin;  
KW antibiotic; biopesticide; virulence factor; disease model; plague;  
KW whooping cough.

XX Photorhabdus luminescens.

XX MO200294867-A2.

XX 28-NOV-2002.

XX 07-FEB-2002; 2002MO-IB003040.

XX 07-FEB-2001; 2001FR-00001659.

PA (INSP ) INST PASTEUR.  
PA (CNRS ) CNRS CENT NAT RECH SCI.

PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A,  
PI Buchrieser C;

XX WPI; 2003-148459/14.

XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
XX

PS Claim 2; SEQ ID NO 379; 1205bp; French.

XX The invention relates to the isolation of genes and their encoded  
CC proteins from Photorhabdus luminescens. The isolated sequences are  
CC sources of probes and primers for detecting the genome of P. luminescens  
CC and related species; to study polymorphisms; for gene analysis and for  
CC detection/amplification of the genes. Antibodies (Ab) raised against the  
CC polypeptides encoded by the genes are used for detection/identification  
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
CC carry a gene-containing vector are used to select compounds that  
CC modulate, regulate, induce or inhibit expression of the genes in plants,  
CC animals or microorganisms other than P. luminescens and are able to alter  
CC response or sensitivity to toxins and antibiotics produced by P.  
CC luminescens. Cells transformed to express the genes are useful for  
CC recombinant production of the proteins, particularly toxins and  
CC antibacterials useful as insecticides, bactericides and fungicides. The  
CC genes, proteins, vectors containing the genes and Ab are also useful  
CC therapeutically (to treat microbial infection by bacteria or fungi that  
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
CC biopesticides. Other uses of the genes and the proteins are as virulence  
CC factors and for identifying targets of human diseases for which P.  
CC luminescens is a model (particularly plague and whooping cough). This  
CC sequence represents one of the isolated P. luminescens proteins  
XX

SQ Sequence 341 AA;

Query Match 8.2%; Score 145.5; DB 6; Length 341;

Best Local Similarity 22.9%; Pred. No. 0.00041; Mismatches 137; Indels 53; Gaps 15;

Matches 74; Conservative 59; Mismatches 137; Indels 53; Gaps 15;

QY 29 NIYANGLMIGVINITPTD-DEGNFVDIDVTLNDNIK-----IVDYIDGSDIDGSD 80  
DB 20 NPYANGHGQCMVAKISVLKQEKYKNGDWYKL--ALSDAEKKSIGVAAALSDLIYDQLKMP 76  
QY 81 GWPYTGPNENY-----NTINSQSIS-----LKSNEQ--ITQIKRYVSC-SNT 122  
DB 77 GWTTPDARNKYDGLGLNGVYSADIFTEEPVIRAGDCCCTENYONSYSVEIILCYVSS 136  
QY 123 SRLRTKSFSAKVT--TTSQVVISITQNSIN-----SRVIVINAIDATNTDDELRTKRETR 176  
DB 137 NRTSTBYLMAKMTPEIDNGKRTLTNNMSVGDEVPDSKILKALAPYANNAQGHESNITL 196  
QY 177 FENQSYTSKRSSTN--SLVYHTWTIPRSLKLOWRMEDYNNNGWTWASQCYKXGADGSGSE 234  
DB 197 FDKTEBPT-KSDTHGQITNLYRWTLPYHLRLLEGNDRTVNNIYVLGRS-----SS 245  
QY 235 STRWLAAGSIFPPGNTDGLWLDNDIALSGMAHKSYNVDGTGINSFTRIIGKGSWYVNI 294  
DB 246 DDRFLTRARVFKGT-SYVARRDMSGGCVWDYSYDVTPDQLAAEVLHVTGGSWTTGY 304  
QY 295 SGLDRGHAVIDQYGN--NKYRI 315  
DB 305 --VDGYHDVTIINDYGCQHKFRI 325

RESULT 3

ABU42557 ID ABU42557 standard; protein; 892 AA.

XX ABU42557;

DT 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #28084.

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Staphylococcus epidermidis.

XX MO200277183-A2.

XX 03-OCT-2002.

XX



```

XX 21-MAR-2002; 2002ZMO-US009107.
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (EILT-) ELITRA PHARM INC.
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
PI Wall D, Travick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX MPI: 2003-0239926/02.
DR N-PSDB; ACAA6427.
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
PS
PS Claim 25; SEQ ID NO 70481; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 623 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 892 AA;
SQ
Query Match 7.1%; Score 126; DB 6; Length 892;
Beat Local Similarity 21.0%; Pred. No. 0.089;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16
QY 9 SDWTGVSFNVILETGLDNCNIVANGLMWG-VIINTPTDDEGNFVDDVT----- 60
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 398 SMFNTIDTKNHTVEQ-----TIYINPLRYSAKETNWNISGNGDBESIT-IDDSIIKIKYK 451
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 61 -----LNDNKKIYDIDGSDIDGSDGKPFYGNPEYNT-----INNS 97
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 452 VGDNONLPDSNRRIYDSYEDVT-NDDYAQACNNNDVAINRGNDSPYIIKIVSKYDPNK 510
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 98 QSAYLSKSENSQITQIKRYVCSNTRSRIRKTSFSAKV-TTTSQ-----KYISI- 144
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 511 DDYTTIQQTVMTQITINYY-----TGEFRIRASVDYMTIAFSSISGGGGDLDPEPKYKIGD 565
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 145 -----TQNSINSSRVAINAIDATNFTDDDELRTTKETRFENQSYTSKHSSTNSI-Y 193

```

Db	566	YVMEVDVDDKGGQNTNNDNEKPLSNVLTULTYPRG---	TSKSVPTDDEGKKQPGFLKNGLTY	622
Qy	194	VHTWTPRSLKLQNMWREDY-----	---NNGWTAOSCYRKTGADGGSSEST	236
Db	623	KITEFTP-----	---EGYTPFLKHSGTNPALDSEGNSSVWVTINGQDDMTIDSGFYQTP	672
Qy	237	RMLAAGSLFFPGNGYDGLMDLDMDIALSGMAHKSYNVDGTGNOISLFRLLIKGSGSWYNIISG		296
Db	673	KYSLGANYWYDYNKDGIGQDDERKISGV--	KVTLDENGNIISTYTTDENG--KYQFDN	727
Qy	297	LDRGHAVIIDIQ	308	
Db	728	LNSGNVYVHFDK	739	
RESULT 4				
ID	AD084849	AD084849 standard; protein; 892 AA.		
XX	AD084849;			
XX	29-JUL-2004	(first entry)		
XX	S	epidermidis surface anchored LPXTG protein SegID21.		
XX	LPXTG;	cell wall-anchored surface protein; Gram positive bacterium;		
XX	extracellular matrix molecule; sequence database; C-terminal;			
XX	immunoglobulin-like fold region; Ig-like fold region; antibacterial;			
XX	vaccine; gene therapy; infection; medical device; prosthesis;			
XX	premature newborn; AIDS; debilitated cancer; bone marrow transplantation.			
XX	Staphylococcus epidermidis.			
XX	MO2004025416-A2.			
XX	25-MAR-2004.			
XX	15-SEP-2003;	2003WO-US028789.		
XX	13-SEP-2002;	2002US-0410303P.		
XX	(TEXA )	UNIV TEXAS A & M SYSTEM.		
XX	(INH1-)	INHIBITEX INC.		
XX	(UABR-)	UAB RES FOUNDD.		
XX	Hook M, Xu Y, Sillanpaa JV, Sthanam N, Ponnuraj K, Patti JM;			
XX	Hutchins JT, Hall A;			
XX	WPI; 2004-315684/29.			
XX	This invention relates to a novel method of identifying LPXTG-containing			
XX	cell wall-anchored surface proteins from Gram positive bacteria that bind			
XX	to an extracellular matrix molecule which comprises searching a database			
XX	of sequence information for a putative protein sequence having the LPXTG-			
XX	motif in its C-terminal region and analysing the sequence for the			
XX	presence of one or more immunoglobulin (Ig)-like fold regions. The			
XX	invention may be useful for the production of compounds with an			
XX	antibacterial activity or for production of a vaccine. In addition the			
XX	disclosed sequences may be useful for gene therapy. The antibody is			
XX	useful for treating or preventing an infection of Gram-positive bacteria			
XX	in a human or animal patient. The method and the proteins are useful in			
XX	generating antibodies for treating and preventing the spread of			
XX	infections of Gram positive bacteria, for interfering with, or inhibiting			
XX	binding interactions by Gram positive bacteria, for monitoring the level			
XX	of gram positive bacterial antigens, or antibodies recognising the			

CC antigens in a human or animal patients suspected of containing the  
CC antigens or antibodies, in preventing or reducing infection of medical  
CC devices and prostheses caused by such organisms, and in treating or  
CC preventing infections in highly susceptible groups such as premature  
CC newborns, AIDS and debilitated cancer patients, and bone marrow  
CC transplantation. The present sequence is that of a surface anchored LPTXG  
CC protein identified using the method of the invention.

XX  
SQ Sequence 892 AA;

Query Match 7.1%; Score 126; DB 8; Length 892;  
Best Local Similarity 21.0%; Pred. No. 0.089;

Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT----- 60  
DB 398 SMFTNIDTKHNTVEQ-----TIIYNPLRYSAKETNVIISGNGDEGSTI-IDDSITIKVYK 451  
QY 61 -----LNDNIKIYDYIDSGDIDSGDMFYTGPNFNEVNT-----IPNS 97  
DB 452 VGDNQNLPSNRRIYDSEYEDVT-NDVYAOLGNNNDVNIINGNDSPYIIKVISKYPNK 510  
QY 98 QSYSLKSENSQITQIKRYVSCNTSRLRTKFSKAKV--TTTSG-----KVISI-- 144  
DB 511 DDYTTIOQTYVMQTINBY-----TGEFRITASYNNTIAFSTSSGCGGDLPEPKTYKIGD 565  
QY 145 -----TQNSINSSRVVINAIDATNFTDDELRTTKETRFENQSYTHKSSTNSL-Y 193  
DB 566 YWEDVDKDGIOQNTNDEKPLSNVLVLTYPDG--TSKSVRTDEBKYPDGLKNGILTY 622  
QY 194 VHTWTIPRSLKLGQWRWEDY-----NNGWTMAQSCYYTKGADGSEST 236  
DB 623 KITETP-----EGYPTLKHSQTNPALDSBGNSVWVTINQDDMTIDSGFYQTP 672  
QY 237 RWLAAGSIFPPGNYDGLMNDIALSGMAKSYVVDGIGINLSFTRIIGKFSWVNIISG 296  
DB 673 KYSLGNYVWDTNKGIGQSDGKGISGV--KVTLKDENGNIISTTTDENG--KYQFDN 727  
QY 297 LDRGHAVIITDQ 308  
DB 728 LNSGNVIYHFDK 739

RESULT 5  
ADS20651  
ID ADS20651 standard; protein; 892 AA.

XX AC ADS20651;

XX DT 30-DEC-2004 (first entry)

XX DE S. epidermidis hyperimmune serum reactive antigen protein - SEQ ID 52.

XX KW antigen; antibiotic resistance; antibacterial; vaccine; gene therapy.

XX OS Staphylococcus epidermidis.

XX PN WO2004087746-A2.

XX 14-OCT-2004.

XX PF 31-MAR-2004; 2004WO-EP003398.

XX PR 31-MAR-2003; 2003EP-00450078.

XX PA (INTE-) INTERCELL AG.

XX PI Meinke A, Min Bui D, Nagy E;

XX DR WPI; 2004-729219/71.

XX DR N-PSDB; ADS20620.

XX PT New nucleic acid molecules encoding hyperimmune serum reactive antigens

PT from Staphylococcus epidermidis, useful for diagnosing, preventing or  
XX treating S. epidermidis infections.

XX PS Claim 11, SEQ ID NO 52; 196pp; English.

XX The invention relates to a novel isolated nucleic acid molecule encoding  
CC a hyperimmune serum reactive antigen or its fragment. Staphylococci are  
CC commonly associated with human disease. Both Staphylococcus epidermidis  
CC and Staphylococcus aureus have become resistant to many commonly used  
CC antibiotics, most importantly methicillin (MRSA) and vancomycin (VISA).  
CC Drug resistance is an increasingly important public health concern and  
CC novel therapies to combat staphylococci infection must be developed in  
CC preparation for a time when such infections may be untreatable by  
CC antibiotics. The molecules of the invention demonstrate antibacterial  
CC activity and may be useful for manufacturing a medicament, such as a  
CC vaccine, for treating or preventing S. epidermidis infections, possibly  
CC via gene therapy. The antigen or its fragment may also be used for  
CC generating an anticalical peptide binding to the hyperimmune serum  
CC reactive antigen or fragment, for manufacturing a functional nucleic  
CC acid, such as an aptamer or spiegelmer and for manufacturing a functional  
CC ribonucleic acid, such as a ribozyme, antisense nucleic acid or siRNA  
CC (short interfering RNA). The current sequence is that of a Staphylococcus  
CC epidermidis hyperimmune serum reactive antigen protein of the invention.

XX  
SQ Sequence 892 AA;

Query Match 7.1%; Score 126; DB 8; Length 892;  
Best Local Similarity 21.0%; Pred. No. 0.089;

Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT----- 60  
DB 398 SMFTNIDTKHNTVEQ-----TIIYNPLRYSAKETNVIISGNGDEGSTI-IDDSITIKVYK 451  
QY 61 -----LNDNIKIYDYIDSGDIDSGDMFYTGPNFNEVNT-----IPNS 97  
DB 452 VGDNQNLPSNRRIYDSEYEDVT-NDVYAOLGNNNDVNIINGNDSPYIIKVISKYPNK 510  
QY 98 QSYSLKSENSQITQIKRYVSCNTSRLRTKFSKAKV--TTTSG-----KVISI-- 144  
DB 511 DDYTTIOQTYVMQTINBY-----TGEFRITASYNNTIAFSTSSGCGGDLPEPKTYKIGD 565  
QY 145 -----TQNSINSSRVVINAIDATNFTDDELRTTKETRFENQSYTHKSSTNSL-Y 193  
DB 566 YWEDVDKDGIOQNTNDEKPLSNVLVLTYPDG--TSKSVRTDEBKYPDGLKNGILTY 622  
QY 194 VHTWTIPRSLKLGQWRWEDY-----NNGWTMAQSCYYTKGADGSEST 236  
DB 623 KITETP-----EGYPTLKHSQTNPALDSBGNSVWVTINQDDMTIDSGFYQTP 672  
QY 237 RWLAAGSIFPPGNYDGLMNDIALSGMAKSYVVDGIGINLSFTRIIGKFSWVNIISG 296  
DB 673 KYSLGNYVWDTNKGIGQSDGKGISGV--KVTLKDENGNIISTTTDENG--KYQFDN 727  
QY 297 LDRGHAVIITDQ 308  
DB 728 LNSGNVIYHFDK 739

RESULT 6  
ABP40469  
ID ABP40469 standard; protein; 930 AA.

XX AC ABP40469;

XX DT 24-JUL-2002 (first entry)

XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5314.

XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX KW antibacterial; gene therapy.

XX OS Staphylococcus epidermidis.

XX US6380370-B1.  
 PN 30-APR-2002.  
 PD 13-AUG-1998; 98US-00134001.  
 PF 14-AUG-1997; 97US-0055779P.  
 PR 08-NOV-1997; 97US-0064964P.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA Doucette-Stamm LA, Bush D;  
 PI WPI; 2002-381255/41.  
 XX N-PSDB; ABN93014.  
 DR Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*  
 XX polypeptide, useful for diagnosing and treating bacterial infections.  
 PT Disclosure; SEQ ID NO 5314; 267pp; English.  
 XX  
 PS ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly *S. epidermidis* infections. The sequences can be used to  
 CC screen for compounds able to interfere with the *S. epidermidis* life cycle  
 CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site  
 XX  
 SQ Sequence 930 AA;

Query Match 7.1%; Score 126; DB 5; Length 930;  
 Best Local Similarity 21.0%; Pred. No. 0.095;  
 Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;  
 QY 9 SDMTGVTSTVAVILETGIDNCNIYANGIMIG--VIINITPTDDEGNFVIDDVT----- 60  
 DB 436 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNNISGNGEGSTI-IDSTIIKVKYK 489  
 QY 61 -----LNDNIKIVDYIDSGSDGWFYTGPNENYNT-----IPNS 97  
 DB 490 VGNQNLPSNRIRYDSEYEDVT-NDYVQLGNNNDVINFGNIDSPYIIKVISKIDPNK 548  
 QY 98 QSYSLKSENSQITQIKRYVSCSWTSRLRTKSFSAKV--TTSG-----KVISI-- 144  
 DB 549 DDVTTTQIYTMGTINNEY-----TGFRTASVYNTIATFSTSSGCGGGLPPEKTYKIGD 603  
 QY 145 -----TQNSINSSRVAINAIDATNFTDELRTTKETREPENOSYTHSKSTNSL-Y 193  
 DB 604 YWEDVDKQDIGNTNDNEKPLSNVLVLTYPDG---TSKSVRDBEKGYPFDGLKNGLTLY 660  
 QY 194 VHWWTIPRSKLQNKWREDEY-----NNGWTAQSYCYTGTADGSEST 236  
 DB 661 KTFETP-----EGYTPILKHSGTNPALDSEGNASVWVITGDDMTIISGFGTPT 710  
 QY 237 RMLAAGSIFFPGYVDGLNDIALSGMAKSYVVDGTIGIQLSFTRIIGKGFVYVNIISG 296  
 DB 711 KYELGNVYVWDTNKDGIQSDGDKGISGV--KVIKLDENGNIITTTTDENG--KIQFDN 765  
 QY 297 LDRGHAVIITDQ 308  
 DB 766 LNSGNVYVHDK 777

RESULT 7  
 ADS06014  
 ID ADS06014 standard; protein; 930 AA.  
 AC ADS06014;

XX 04-NOV-2004 (first entry)  
 DT *Staphylococcus epidermidis* polypeptide seqid 5309.  
 DE antibacterial; vaccine; antisense therapy; *Staphylococcus epidermidis*;  
 XX recombinant expression vector; infection; computer readable medium;  
 XX computer based system.  
 OS *Staphylococcus epidermidis*;  
 XX US2004147734-A1.  
 PN 29-JUL-2004.  
 XX 01-DEC-2003; 2003US-00724972.  
 PF 08-NOV-1997; 97US-0064964P.  
 PR 13-AUG-1998; 98US-00134001.  
 PR 29-NOV-1999; 99US-00450969.  
 XX (DOUC/) DOUCETTE-STAMM L.  
 PA (BUSH/) BUSH D.  
 PI Doucette-Stamm L, Bush D;  
 DR WPI; 2004-580138/56.  
 XX N-PSDB; ADS02242.  
 DR  
 XX New isolated polypeptide and encoding nucleic acid derived from  
 PT *Staphylococcus epidermidis*, useful for diagnosing, preventing and/or  
 PT treating an *S. epidermidis* bacterial infection.  
 XX  
 XX Claim 17; SEQ ID NO 5309; 741pp; English.  
 PS  
 XX The invention describes an isolated nucleic acid comprising a nucleotide  
 CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
 CC 1-3772) and encoding an *Staphylococcus epidermidis* polypeptide with any  
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
 CC given in the specification. Also described are: a recombinant expression  
 CC vector; a cell comprising a recombinant expression vector of (1);  
 CC producing an *S. epidermidis* polypeptide; an isolated nucleic acid  
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
 CC vaccine composition for prevention or treatment of an *S. epidermidis*;  
 CC infection, comprising a nucleic acid cited above and a carrier; treating  
 CC a subject for *S. epidermidis* infection; a recombinant or substantially  
 CC pure preparation of an *S. epidermidis* polypeptide or its fragment; a  
 CC vaccine composition for prevention or treatment of an *S. epidermidis*;  
 CC infection; detecting the presence of a *Staphylococcus* nucleic acid in a  
 CC sample; a computer readable medium having recorded in it the nucleotide  
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
 CC system for identifying fragments of the *Staphylococcus* genome of  
 CC commercial importance; a computer based system for identifying fragments  
 CC of the *Staphylococcus* plasmids of commercial importance; identifying  
 CC commercially important nucleic acid fragments of the *Staphylococcus*  
 CC genome and/or plasmids; and identifying an expression modulating fragment  
 CC of the *Staphylococcus* genome and/or plasmids. The methods and  
 CC compositions of the present invention are useful for the diagnosis,  
 CC prevention and/or treatment of an *Staphylococcus epidermidis* bacterial  
 CC infection. This is the amino acid sequence of a *S. epidermidis* protein of  
 CC the invention.  
 XX  
 SQ Sequence 930 AA;

Query Match 7.1%; Score 126; DB 8; Length 930;  
 Best Local Similarity 21.0%; Pred. No. 0.095;  
 Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;  
 QY 9 SDMTGVTSTVAVILETGIDNCNIYANGIMIG--VIINITPTDDEGNFVIDDVT----- 60  
 DB 436 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNNISGNGEGSTI-IDSTIIKVKYK 489  
 QY 61 -----LNDNIKIVDYIDSGSDGWFYTGPNENYNT-----IPNS 97

```

Db      490 VGDNQNLPSDNRIYDSEYEDVT--NDDYAQLGNNNDVNIINGNIDSPYIIKVISKYPNK 548
Qy      98 QSVSLKSENSQITQIKRYVSCNTSLRTRKSPSAKY--TTTSG-----KVISI-- 144
Db      549 DDYTTIQVTVMQTTINEY-----TGEFRITASYDNTIAFTSSQGGGDLPEPRTYKIGD 603
Qy      145 -----TONSINSSRVVINAIDNTNFTDDELRTTKETREPNQSYTHKSTNSL-Y 193
Db      604 YWEDVDVKDGIQTNDNEKELSNVLVLTYPDG---TSKSVRTDEBKQYDGLKNGLTY 660
Qy      194 VHTWTIPRSLKLQWNRMEDY-----NNGWTMAQSCYYTGTADGSEST 236
Db      661 KITFETP-----EGYTPPLKHSGTNPALDSEGNVWVTINGDDMTIDSGFYQTP 710
Qy      237 RMLAAGSIFPPGNVDGLMLNDIALSGMAHKSVMVDGINGDLSPTRIIGKFSWVYNIISG 296
Db      711 KYSLGNYWYDYNKDGIGDDEKGISGV--KVTLKDENGNIISTTTDENG--KYQPDN 765
Qy      297 LDRGHAVIIDIQ 308
Db      766 LNSGNVIVHFDK 777

```

RESULT 8  
ID AAV83171 standard; protein: 991 AA.

AAV83171;

24-JUN-2000 (first entry)

Cell wall protein Sdrg.

Sdrg; Sdrg; Sdrg; coagulase negative; staphylococcus; septicemia;  
osteomyelitis; endocarditis; immune response; vaccine; graft; stent;  
intravenous catheter; heart valve; cardiac.

Staphylococcus sp.

Key Location/Qualifiers

Misc-difference 14 /note= "Position encoded by TAG stop codon"

Misc-difference 33 /note= "Position encoded by TGA stop codon"

Misc-difference 964 /note= "Position encoded by TAA stop codon"

Misc-difference 980 /note= "Position encoded by TAG stop codon"

Misc-difference 989 /note= "Position encoded by TAA stop codon"

WO200012689-A1.

09-MAR-2000.

31-AUG-1999; 99WO-US019728.

31-AUG-1998; 98US-0098443P.

25-JAN-1999; 99US-0117119P.

(QUEE-) QUEEN ELIZABETH COLLEGE DUBLIN.

(TEXA) UNIV TEXAS A & M SYSTEM.

Foster TJ, Hook M, Davis S, Hartford O, McCreia K, Ni Eidhin D;

WPI, 2000-256637/22.

N-PSDB; AA293534.

Recombinant or synthetic proteins from coagulase-negative staphylococci  
useful for prevention, treatment and diagnosis of staphylococcal  
infections bind soluble and immobilized fibrinogen.

PS Claim 8; Fig 3; 104pp; English.  
XX Isolated staphylococcus sdr cell wall proteins which bind both soluble  
CC and immobilized fibrinogen are useful for treating or preventing  
CC coagulase-negative staphylococcal infection such as septicemia,  
CC osteomyelitis or endocarditis, and for inducing immune responses in  
CC patients. The cell wall proteins are also useful for reducing coagulase-  
CC negative staphylococcal infection of indwelling medical devices such as  
CC vascular grafts, vascular stents, intravenous catheters, artificial heart  
CC valves and cardiac assist devices. The cell wall associated proteins are  
CC able to inhibit staphylococcal adhesion to immobilised extracellular  
CC matrix or host cells present on the surface of implanted biomaterials  
XX

Sequence 991 AA;

Query Match 7.1%; Score 126; DB 3; Length 991;  
Best Local Similarity 21.0%; Pred. No. 0.1; Indels 102; Gaps 16;  
Matches 78; Conservative 51; Mismatches

Qy 9 SDWTGVSTFNVILETGLDNGNIYANGIMIG--VIINIFPTDDEGNFVDIDVT----- 60  
Db 469 SMFTNIDTKHMYEQ-----TIYINPLRYSAKETNNVNSGNGEGSTI-IDSTIIIVYK 522

Qy 61 -----LNDNIKIVDYIDSDIDSGDGEYTGPNPEYNT-----IPNS 97

Db 523 VGDNQNLPSDNRIYDSEYEDVT--NDDYAQLGNNNDVNIINGNIDSPYIIKVISKYPNK 581

Qy 98 QSVSLKSENSQITQIKRYVSCNTSLRTRKSPSAKY--TTTSG-----KVISI-- 144

Db 582 DDYTTIQVTVMQTTINEY-----TGEFRITASYDNTIAFTSSQGGGDLPEPRTYKIGD 636

Qy 145 -----TONSINSSRVVINAIDNTNFTDDELRTTKETREPNQSYTHKSTNSL-Y 193

Db 637 YWEDVDVKDGIQTNDNEKELSNVLVLTYPDG---TSKSVRTDEBKQYDGLKNGLTY 693

Qy 194 VHTWTIPRSLKLQWNRMEDY-----NNGWTMAQSCYYTGTADGSEST 236

Db 694 KITFETP-----EGYTPPLKHSGTNPALDSEGNVWVTINGDDMTIDSGFYQTP 743

Qy 237 RMLAAGSIFPPGNVDGLMLNDIALSGMAHKSVMVDGINGDLSPTRIIGKFSWVYNIISG 296

Db 744 KYSLGNYWYDYNKDGIGDDEKGISGV--KVTLKDENGNIISTTTDENG--KYQPDN 798

Qy 297 LDRGHAVIIDIQ 308  
Db 799 LNSGNVIVHFDK 810

RESULT 9  
ID AAV70120 standard; protein: 991 AA.

AAV70120;

06-JUN-2000 (first entry)

Staph. epidermidis serine-aspartate repeat region protein Sdrg.

Multi-component vaccine; immunostimulatory; antibacterial; MSCRAMM;  
microbial surface components recognising adhesive matrix molecules;  
collagen binding protein; CBP; CNA; fibrinogen binding protein;  
clumping factor A; ClfA; Clumping factor B; ClfB; Fbp;  
fibrinectin binding protein; Staphylococcus infection;  
serine-aspartate repeat region protein; SDR protein; Sdrg.

Staphylococcus epidermidis.

Key Location/Qualifiers

Misc-difference 14 /note= "Encoded by in-frame stop codon TAG"

Misc-difference 33 /note= "Encoded by in-frame stop codon TGA"

Misc-difference 964

```

FT /note= "Encoded by in-frame stop codon TAA"
FT Misc-difference 980
FT /note= "Encoded by in-frame stop codon TAG"
FT Misc-difference 989
FT /note= "Encoded by in-frame stop codon TAA"
XX
XX MO200012131-A1.
XX
XX 09-MAR-2000.
XX
XX 31-AUG-1999; 99WC-US019727.
XX
XX 31-AUG-1998; 98US-0098439P.
XX
XX (INHIT-) INHIBITEX INC.
XX (TEXA-) UNIV TEXAS A & M SYSTEM.
XX (QUEE-) QUEEN ELIZABETH COLLEGE DUBLIN.
XX
XX Patti JM, Foster TJ, Hook M;
XX
XX WPI; 2000-237781/20.
XX N-PSDB; AA251202.
XX
XX Composition used for generating immune response or for inhibiting
XX microbial colonization in an animal comprises antibodies that bind
XX collagen binding protein, fibrinogen binding protein and, optionally,
XX fibrinectin binding protein.
XX
XX Claim 8; Fig 4; 115pp; English.
XX
XX The patent discloses multicomponent vaccines containing selected
XX combinations of bacterial binding proteins termed MCRAMM (microbial
XX surface components recognising adhesive matrix molecules) or their
XX antibodies. A vaccine composition is provided that includes collagen
XX binding protein or peptide, e.g. CNA, a fibrinogen binding protein
XX preferably Clumping factor A (ClfA) or Clumping factor B (ClfB), and
XX optionally a fibrinectin binding protein e.g. FnBP-A. The vaccines are
XX useful for imparting protection against a broad spectrum of
XX Staphylococcal strains and for inhibiting microbial colonization,
XX especially of Staphylococcus aureus, in an animal. The combinations can
XX also be used to select donor blood pools for the preparation of purified
XX blood products for passive immunisation. The present sequence is a serine
XX -aspartate repeat region protein, Sdrg from Staphylococcus epidermidis.
XX The Sdr protein is useful in vaccine preparation in combination with
XX specific bacterial binding proteins. These vaccines can be used to treat
XX a broad spectrum of bacterial infections, including those arising from
XX both coagulase-positive and coagulase-negative bacteria
XX
XX Sequence 991 AA:
XX
XX Query Match 7.1%; Score 126; DB 3; Length 991;
XX Best Local Similarity 21.0%; Pred. No. 0.1;
XX Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;
XX
XX 9 SDWTGTVFVILETGLDNCNIYANGLMIG--VIINIFPTDDEGNFVDIDVT----- 60
XX | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX SMTETNIDTKHYTEQ-----TIYNIPRYSAKETNVIISNGSGSIT-IDDSITIKVVK 522
XX
XX 61 -----LNDNIKIVDYIDGSDGWFYTGPNENYNT-----IPNS 97
XX | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX VCDNQMLPDSNRIVDYSEVEDVT--NDVYAQLGNNNDVINFGNIDSPYIIKIVSKYDPNK 561
XX
XX 98 QSVSLKSEKSOITQIKRYVSCNSTRKTSFSAKV--TTTG-----KVSI-- 144
XX | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX DDVTITQQTVMQTTINEY-----TGEFRYASVDNTIAFTSSGGQGGDLPPERTYKIGD 636
XX
XX 145 -----TQNSINSSRVVINAIDATNFTDELRTTKETRENOGYSHKSTNSL--Y 193
XX | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX YWVEDVDKQIQNTNNEKEPLSNVLYLVYTPDG---TSKSVKRDDEBKVPDGLKNGLTY 693
XX
XX 194 VHTWTTPRSKLKLNWRMEDY-----NNGWTVAQSCYRYTGADGSEST 236
XX | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX KITFETP-----EGYTPTLKISGTNPALDSEGNVWVTINGDDWTIDSGFYQTP 743
XX

```

```

QY 237 RMLAGSIPPPGNYDGLMDNDIALSGMAKSYNVDITGINSFTRIIGKPSWYNISG 296
DB 744 KSLGNYWYDNYKDDIQGDEKISGV--KVTLKQENGNITSTTTDENG---KYQFDN 798
QY 297 LDRGHAVIIDQ 308
DB 799 LNSGNYIVHFDK 810
XX
XX RESULT 10
XX AA#41602
XX ID AA#41602 standard; protein; 1092 AA.
XX
XX AA#41602;
XX
XX 17-OCT-2003 (revised)
XX 22-JUN-1998 (first entry)
XX
XX Staphylococcus epidermidis fibrinogen binding protein FIG.
XX
XX Fibrinogen binding protein; FIG; aggregation; infection;
XX coagulase-negative Staphylococcus; therapy; diagnosis; immunisation;
XX immunogen; vaccine.
XX
XX Staphylococcus epidermidis; strain HB.
XX
XX Key Location/Qualifiers
XX Peptide 1..51
XX Protein /label= Sig_peptide
XX Region /label= Mat_protein
XX Region 52..824
XX Region /note= "non-repetitive region, harbours fibrinogen
XX binding activity"
XX Region 825..1040
XX Region /note= "Asp-Ser dipeptide repeat region"
XX Region 1053..1057
XX Region /note= "cell wall anchoring motif"
XX
XX MO9748727-A1.
XX
XX 24-DEC-1997.
XX
XX 18-JUN-1997; 97WC-SE001091.
XX
XX 20-JUN-1996; 96SE-00002496.
XX
XX (GUSG/) GUSG B.
XX (NILS/) NILSSON M.
XX (FRYK/) FRYKBERG L.
XX (FLOC/) FLOCK J.
XX (LIND/) LINDBERG M.
XX
XX Guss B, Nilsson M, Frykberg L, Flock J, Lindberg M;
XX WPI; 1998-063079/06.
XX N-PSDB; AA#04279.
XX
XX Fibrinogen-binding protein from coagulase-negative Staphylococcus - used
XX for prevention, treatment and diagnosis of Staphylococcus infection.
XX
XX Example 3; Fig 6; 45pp; English.
XX
XX The protein comprises the fibrinogen binding protein (FIG) of coagulase-
XX negative Staphylococcus epidermidis HB. Its amino acid sequence was
XX deduced from the isolated fig gene (see AA#04279). The closest known
XX analogue of FIG is the clumping factor of S. aureus which also binds
XX fibrinogen and promotes bacterial aggregation in serum. Recombinant FIG
XX polypeptides can be expressed in host cells. They are used as immunogens,
XX particularly in vaccines (which may be expressed in vivo) to protect
XX humans and animals against coagulase-negative Staphylococcus infection.
XX Antibodies raised against FIG can be used for passive immunisation. They
XX

```

CC block the adherence of bacteria) and for diagnosis. (Updated on 17-OCT-  
 CC 2003 to standardise OS field)  
 XX  
 SQ Sequence 1092 AA;

Query Match 7.0%; Score 124; DB 2; Length 1092;  
 Best Local Similarity 21.0%; Pred. No. 0.18;  
 Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

QY 9 SDMTGVSTFVILETGIDNCNIYANGIMIG--VIINITEPDEGNFVDIDVT-----60  
 DB 439 SMFTNIDTKNHYEQ-----TIYINPLRSAKETNNVNSGNGEGSTI-IDDSTIIKVKY 492  
 QY 61 -----LNDNIKIVDYIDSGDIDSDGMPYTGPNENEYNT-----IPNS 97  
 DB 493 VGDNQNLPSNRIVDYSEYEDVT-NDYVQLGNNNDVINFGNIDSPYIIKVISKYDPNK 551  
 QY 98 QSYSLKSENSQITQIKRYVSCSNTSRLRTKFSKAV--TTTSG-----KVISI--144  
 DB 552 DDYTTIQQVYTMQTTINEY-----TGEFRFASVDNTIAFSTSGQGQGDLPPEKTYKIGD 606  
 QY 145 -----TONSINSRVVINAIDATNFTDDELRTTKETRFENOSYTSKSTNSL-Y 193  
 DB 607 YWEDVDKDGIQNTDNEKPLSNVLTLYTPDG---TSKSVRTDEDEKQYFDGLKNGLTY 663  
 QY 194 VHTWTIPRSLKIQWREDEY-----NNGWTWAGSCYVYTGADGSGEST 236  
 DB 664 KITFETP-----EGYTPILKHSCTNPALDSEGNVWVTINQDDMTIDSGFYQTP 713  
 QY 237 RMLAAGIFPPGNYDGLMNDIALSGMAHKSYNVDGINGQLSPTRIIGKFSWVNI 296  
 DB 714 KYSLGNYVWYDNTKDGIQGDEKGISGV--KVTLKDENGNIISTTTDENG---KYQFDN 768  
 QY 297 LDRGHAVIITDQ 308  
 DB 769 LNSGNIVVHFDK 780

## RESULT 11

ABM79019  
 ID ABM79019 standard; protein; 1092 AA.

XX ABM79019;  
 AC  
 XX 15-JAN-2004 (first entry)  
 DT  
 XX Staphylococcus epidermidis polypeptide.  
 DE  
 XX Infection; antibacterial; vaccine.  
 KW  
 XX Staphylococcus epidermidis.  
 OS  
 XX WO2003076470-A1.  
 PN  
 XX 18-SEP-2003.  
 PD  
 XX 05-MAR-2003; 2003WO-US006415.  
 PE  
 XX 05-MAR-2002; 2002US-0361324P.  
 PR  
 XX (INH1-) INHIBITEX INC.  
 PA (TEXA ) UNIV TEXAS A & M SYSTEM.  
 XX  
 PI Patti JM, Hutchins JT, Hall A, Domanski P, Patel P, Hook M,  
 FI Robbins J, Vernachio J, Bowden MG;  
 XX  
 DR WPI; 2003-722324/68.  
 XX  
 PT New antibody recognizing a Staphylococcus epidermidis protein comprising  
 PT SdGc NIN23, SdGc N2N3 or SdGcR2 useful for preparing a composition for  
 FT treating or preventing a coagulase-negative Staphylococcal infection.  
 XX  
 PS Claim 16; Page 36-37; 78pp; English.

XX  
 CC The present sequence comprises the protein sequence of a polypeptide of a  
 CC coagulase-negative Staphylococcus epidermidis. A claimed monoclonal  
 CC antibody recognises this protein and is used in a claimed method of  
 CC treating or preventing a coagulase-negative staphylococcal infection in a  
 CC human or animal, e.g. a nosocomial coagulase-negative staphylococcal  
 CC infection in low birth weight infants  
 XX  
 SQ Sequence 1092 AA;

Query Match 7.0%; Score 124; DB 7; Length 1092;  
 Best Local Similarity 21.0%; Pred. No. 0.18;  
 Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

QY 9 SDMTGVSTFVILETGIDNCNIYANGIMIG--VIINITEPDEGNFVDIDVT-----60  
 DB 439 SMFTNIDTKNHYEQ-----TIYINPLRSAKETNNVNSGNGEGSTI-IDDSTIIKVKY 492  
 QY 61 -----LNDNIKIVDYIDSGDIDSDGMPYTGPNENEYNT-----IPNS 97  
 DB 493 VGDNQNLPSNRIVDYSEYEDVT-NDYVQLGNNNDVINFGNIDSPYIIKVISKYDPNK 551  
 QY 98 QSYSLKSENSQITQIKRYVSCSNTSRLRTKFSKAV--TTTSG-----KVISI--144  
 DB 552 DDYTTIQQVYTMQTTINEY-----TGEFRFASVDNTIAFSTSGQGQGDLPPEKTYKIGD 606  
 QY 145 -----TONSINSRVVINAIDATNFTDDELRTTKETRFENOSYTSKSTNSL-Y 193  
 DB 607 YWEDVDKDGIQNTDNEKPLSNVLTLYTPDG---TSKSVRTDEDEKQYFDGLKNGLTY 663  
 QY 194 VHTWTIPRSLKIQWREDEY-----NNGWTWAGSCYVYTGADGSGEST 236  
 DB 664 KITFETP-----EGYTPILKHSCTNPALDSEGNVWVTINQDDMTIDSGFYQTP 713  
 QY 237 RMLAAGIFPPGNYDGLMNDIALSGMAHKSYNVDGINGQLSPTRIIGKFSWVNI 296  
 DB 714 KYSLGNYVWYDNTKDGIQGDEKGISGV--KVTLKDENGNIISTTTDENG---KYQFDN 768  
 QY 297 LDRGHAVIITDQ 308  
 DB 769 LNSGNIVVHFDK 780

## RESULT 12

ADV16755  
 ID ADV16755 standard; protein; 893 AA.

XX ADV16755;  
 AC  
 XX 24-FEB-2005 (first entry)  
 DT  
 XX E. faecalis V583 hyperimmune serum reactive antigen protein - SEQ ID 450.  
 DE  
 XX antigen; antibacterial; vaccine; enterococcus infection; infection;  
 KW pharyngitis; impetigo; rheumatic fever; antipyretic; antirheumatic;  
 KW immunosuppressive; sepsis.  
 OS  
 XX Enterococcus faecalis V583.  
 XX  
 FH Location/Qualifiers  
 FT 6..30  
 FT /note= "Serum reactive epitope, claimed"  
 FT 36..42  
 FT /note= "Serum reactive epitope, claimed"  
 FT 143..157  
 FT /note= "Serum reactive epitope, claimed"  
 FT 176..197  
 FT /note= "Serum reactive epitope, claimed"  
 FT 202..209  
 FT /note= "Serum reactive epitope, claimed"  
 FT 216..233  
 FT /note= "Serum reactive epitope, claimed"  
 FT 226..269  
 FT Region

[illegible]

XX	The invention relates to a novel isolated nucleic acid molecule encoding
CC	a hyperimmune serum reactive antigen e.g. SEQ ID 171-340, 357-372 or 425-
CC	476, or its fragment. The antigen of the invention demonstrates
CC	antibacterial activities and may be useful for producing a pharmaceutical
CC	preparation, particularly a vaccine, against enterococcal infection.
CC	Conditions associated with bacterial infection which may be prevented or
CC	treated include bacterial pharyngitis, scarlet fever, impetigo, rheumatic
CC	fever, necrotizing fasciitis and sepsis in humans. The current sequence
CC	is that of an Enterococcus faecalis V583 hyperimmune serum reactive
CC	antigen protein of the invention.
XX	
SQ	Sequence 893 AA;
Query Match	6.7%; Score 119; DB 9; Length 893;
Best Local Similarity	19.4%; Pred. No. 0.38;
Matches	74; Conservative 58; Mismatches 119; Indels 130; Gaps 20
OY	21 LETGLDN-----CNVANGLNMGVIINITPTDDGCFVDID---DVTINDNKKIVD 69
DB	138 MKNGISSRVARATVAANYANDPNLPG-----KNFIDVSSWNGDISVAEYOIKIS 186
OY	70 Y-IQSGSDIDSGDPFYTCGNPNEBNTVPMSQSVSLKSENSQTQTKRYVSCSNTSRLLTK 128
DB	187 YGVTVGSVKLTREGTWY-NPYAAGQIRNAKAAGLKVS---AHYSMTVSA-A-TAODEAR 240
OY	129 SPSAKVTTSQKVISTONSINSRVINADATNFDTDELRTKETRFENOSYTHSKSS 188
DB	241 YF-AQAANSG-----LDKYTIMNDADPPLTNGGNAAHANSVAFNQOLKALGY 290
OY	189 TNSLYVHTWT-----PRSIKONRWEDYNNGWTMAQCY 225
DB	291 NDALYYGKMVLTTNGYIDTSAFGRDRVMWAQYPYTPDS---SMQWNNDHGAMQWSSQMVF 346
OY	226 KTGADGSESTRMLAAGIFPPRG--NYDGMLUNDNDIALSGMAH-----KEY----- 269
DB	347 -----PGLANYGRRPFDISMYSNFLMNGNSSGPDLISKTYTNP 385
OY	270 -----NVDTGINOLF-----TRIIGKGFSWYNIISGLDRGAHVIIIDO---YG 310
DB	386 GRVIMKNDITFYQDAFRFTPCWRVKAKTLVITIKGLE--YSSAGIQR---LVTDGGYLTA 439
OY	311 NKYRIFLPAGYENSDFYLSSS 331
DB	440 NKDYVL--AAQSNIIDLFFETTN 458
RESULT 13	
ID	ADH86812
AC	ADH86812 standard; protein; 899 AA.
XX	
XX	ADH86812;
DT	22-APR-2004 (first entry)
DE	Enterococcus faecalis polypeptide #1292.
XX	
KM	Enterococcus faecalis infection; transcription regulatory element;
XX	antibacterial.
OS	Enterococcus faecalis.
PN	US6617156-B1.
PD	09-SEP-2003.
PE	13-AUG-1998; 98US-00134000.
PR	15-AUG-1997; 97US-0055778P.
PA	(DOUC/) DOUCETTE-STAMM L A.
XX	(BUSH/) BUSH D.
XX	







Db 115 YTTIQVTVMQNTINEYTFEASVNDTIAFTSSGQGGDLPEKTYKIGPVWEDVDKDG 174  
 Qy 145 TQNSINSRVVINAIDATNFTDDELRTTKETREPNOSYTSKSTNSL-YVHTTIRSL 203  
 Db 175 IQNTNDEKLSNVLTLYTPDG---TSKSVRTDEDEKQYFDGKMLTKYTFEPT--- 228  
 Qy 204 KLGWNRWEDY-----NNGWTMAQCYCYVTGADQGESSTRWLAAGSIFP 246  
 Db 229 -----EGYPTPLKSGTNPALDSEGNVWVTINGDDMTIDSGFYQTKYSLGNYWY 281  
 Qy 247 PGVYDGLMDNDIALSGMAKSYNVDGTINQLSFTRIIGKGFVWYNIISGLDRGHAVIIT 306  
 Db 282 DTWMDGIGDDEKISGV--KVTLKDENGNIISTTTDENG---KYQPDNLNGNYIVHF 336  
 Qy 307 DQ 308  
 Db 337 DK 338

## RESULT 15

AEB91631  
 ID AEB91631 standard; protein; 773 AA.

AC AEB91631;  
 XX 20-OCT-2005 (first entry)

DE Microbial pathogen adhesin protein sequence, SEQ ID NO:341.

XX algorithm; adhesin; pharmaceutical; vaccine; drug screening;  
 KM bordetella pertussis infection; antibacterial; pneumonia;  
 KM antiinflammatory; respiratory-gen.; gastric ulcer; anticancer;  
 KM gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.

OS Shigella flexneri.

PN MO2005076010-A2.

XX 18-AUG-2005.

PF 07-FEB-2005; 2005WO-IN000037.

PR 06-FEB-2004; 2004IN-DB000173.

PR 20-JUL-2004; 2004US-0589227P.

XX (COUL ) COUNCIL SCT & IND RBS SOUTH AFRICA.

PA Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramchandran S;

PI WPI; 2005-597835/61.

PT Computational method for identifying adhesin and adhesin like molecules,  
 PT comprises computing sequence-based attributes of protein sequences using  
 PT neural network software and training an artificial neural network.

PS Claim 17, SEQ ID NO 341; 402pp; English.

CC The present invention relates to a computational method (M1) for  
 CC identifying adhesin and adhesin-like proteins, by computing the sequence-  
 CC based attributes of protein sequences using five attribute modules of a  
 CC neural network software, training an artificial neural network (ANN) for  
 CC each of the computed five attributes, and identifying the adhesin and  
 CC adhesin-like proteins having probability of being an adhesin (Pad) as  
 CC equal or greater than 0.51. Also claimed is a set of 274 annotated genes  
 CC encoding adhesin and adhesin-like proteins, having 274 fully defined 162-  
 CC 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical  
 CC genes encoding adhesin and adhesin-like proteins, having 105 fully  
 CC defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated  
 CC adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base  
 CC pair (SEQ ID NO: 1-278); a set of 105 hypothetical adhesin and adhesin-  
 CC like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:  
 CC 280-384) sequences; and a fully connected multilayer feed forward ANN (I)  
 CC based on (M1). (M1) is useful for identifying adhesin and adhesin-like

CC proteins, of therapeutic potential, and identifying and short-listing  
 CC proteins for further testing in development of new vaccine formulations  
 CC to eliminate diseases caused by various pathogenic organisms. (M1) is  
 CC useful for identifying putative adhesins that are important in drug  
 CC discovery and preventing therapeutics for whooping cough, pneumonia,  
 CC gastric ulcer and urinary tract infections. (M1) identifies adhesins from  
 CC distantly related organisms, and from bacteria belonging to a wide  
 CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of  
 CC unique proteins. The present sequence is a microbial pathogen adhesin  
 CC protein sequence.

SQ Sequence 773 AA;

Query Match 6.6%; Score 118; DB 9; Length 773;

Best Local Similarity 22.7%; Pred. No. 0.37;

Matches 83; Conservative 55; Mismatches 151; Indels 76; Gaps 18;

Qy 2 TLREDSMDWTGVSTFNVILET-----GLDNCNIYANGNMGVIINITPTDDEG 51  
 Db 76 SLDEESTAD-TGSNNENNAIAKIDMAGEITTHGTESYAAYAN-----GTIVKAGDTLDYT 129  
 Qy 52 N---FVDIDVTLDNLIKIDYIDGSDIDGDMFYTGPNF-----YN---TIPNSQ 98  
 Db 130 NASVTLDVDITTHGNNAHAIAARQGT-VSFNGEIIYTTGPDAAIAKIYNGTIVTLKNTS 188  
 Qy 99 SYS-----LKSEBSQITQIRRYVSCSTSLRT-----KPSAKVTTSGKVIAT 144  
 Db 189 AVAHQSGGLVLESSINGQ---SATVDILSGSLRSANIELYHKDETSNVTITDSEVSA 244  
 Qy 145 TQNSINSRVVINAIDATNFTDDELRTTKETREPNOSYTSKSTNSLYVHTTIRSLK 204  
 Db 245 ADVFINNIKHL-TVDATN---SKITGSANISTDDVHTYLSLSDS---TWDIADST 296  
 Qy 205 LQNRWEDYNNGTMAQSCYCYTGADGS-ESTRWLAAGSIFPPGYDGLMDNDIALSG 263  
 Db 297 VSN-----LTVNSFTVIYISRADRDVEPRLTITENYV--GNNGLVHLTELDNDDN 345  
 Qy 264 MAKSYNVDGTINQLSFTRIIGKGFVWYNIISGLDRGHAVIITDQYN---KYRLIFHA 319  
 Db 346 SATDKVINGNTSGTTRKVTNAGSGAATLNGIE---IISVEGSGNBEFTDSTRIFAG 401  
 Qy 320 GYENS 324  
 Db 402 AYEYS 406

Search completed: January 30, 2006, 09:46:03  
 Job time : 28.6283 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2006, 09:42:59 ; Search time 7.48879 Seconds  
(without alignments)  
3687.335 Million cell updates/sec

Title: US-09-889-874a-22

Perfect score: 1784  
Sequence: 1 FTLRSDMSDMWTGVSTFNV.....ILFHAGYNSDPYLSSTIVY 334

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodaca/1/1aa/5 COMB.pep.\*  
2: /cgn2\_6/ptodaca/1/1aa/6 COMB.pep.\*  
3: /cgn2\_6/ptodaca/1/1aa/H.COMB.pep.\*  
4: /cgn2\_6/ptodaca/1/1aa/PTUS.COMB.pep.\*  
5: /cgn2\_6/ptodaca/1/1aa/RB.COMB.pep.\*  
6: /cgn2\_6/ptodaca/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126	7.1	930	US-09-134-001C-5314	Sequence 5314, Ap
2	126	7.1	930	US-09-386-962C-10	Sequence 10, Appl
3	124	7.0	1092	US-09-147-405B-15	Sequence 15, Appl
4	119	6.7	899	US-09-134-000C-4697	Sequence 4697, Ap
5	115.5	6.5	1411	US-10-080-505-17	Sequence 17, Appl
6	114	6.4	360	US-09-107-433-5178	Sequence 5178, Ap
7	114	6.4	376	US-09-583-110-5138	Sequence 5138, Ap
8	112.5	6.3	1436	US-10-080-505-13	Sequence 13, Appl
9	107.5	6.0	334	US-09-830-230A-418	Sequence 418, App
10	107.5	6.0	343	US-09-830-230A-417	Sequence 417, App
11	106	5.9	258	US-09-251-645-5	Sequence 5, Appl1
12	105	5.9	1430	US-09-008-172-2	Sequence 2, Appl1
13	105	5.9	1430	US-09-210-361-6	Sequence 6, Appl1
14	105	5.9	1430	US-09-740-274-6	Sequence 6, Appl1
15	104	5.8	1306	US-09-538-092-330	Sequence 330, App
16	104	5.8	1589	US-09-543-681A-4998	Sequence 4998, Ap
17	103.5	5.8	582	US-09-147-405B-13	Sequence 13, Appl
18	103.5	5.8	583	US-09-147-405B-11	Sequence 11, Appl
19	103	5.8	494	US-09-543-681A-4773	Sequence 4773, Ap
20	103	5.8	607	US-08-409-995-6	Sequence 6, Appl1
21	103	5.8	607	US-08-685-467-6	Sequence 6, Appl1
22	103	5.8	607	US-08-913-942-6	Sequence 6, Appl1
23	103	5.8	607	US-09-684-707-6	Sequence 6, Appl1
24	103	5.8	1395	US-10-080-505-7	Sequence 7, Appl1
25	103	5.8	1912	US-08-409-995-4	Sequence 4, Appl1
26	103	5.8	1912	US-08-685-467-4	Sequence 4, Appl1
27	103	5.8	2353	US-09-377-155-33	Sequence 33, Appl

28	103	5.8	2353	2	US-08-913-942-4	Sequence 4, Appl1
29	103	5.8	2353	2	US-09-669-974-33	Sequence 33, Appl
30	103	5.8	2353	2	US-09-797-862-33	Sequence 33, Appl
31	103	5.8	2353	2	US-09-684-707-4	Sequence 4, Appl1
32	103	5.8	2354	2	US-09-268-347-47	Sequence 47, Appl
33	103	5.8	2411	2	US-09-268-347-36	Sequence 36, Appl
34	102.5	5.7	1963	2	US-09-583-110-5243	Sequence 5243, Ap
35	102.5	5.7	1967	2	US-09-107-433-4883	Sequence 4883, Ap
36	101.5	5.7	545	2	US-09-604-957-4	Sequence 4, Appl1
37	101.5	5.7	545	2	US-09-995-749A-10	Sequence 10, Appl
38	101.5	5.7	1224	2	US-09-901-572A-4	Sequence 4, Appl1
39	101	5.7	1434	2	US-10-080-505-9	Sequence 9, Appl1
40	101	5.7	1457	2	US-09-673-896-4	Sequence 4, Appl1
41	101	5.7	1457	2	US-09-303-518D-650	Sequence 650, App
42	100.5	5.6	829	2	US-09-514-599-6	Sequence 6, Appl1
43	100.5	5.6	829	2	US-09-996-024-6	Sequence 6, Appl1
44	100.5	5.6	862	2	US-09-346-237-1	Sequence 1, Appl1
45	100.5	5.6	1781	2	US-09-395-749A-2	Sequence 2, Appl1

## ALIGNMENTS

RESULT 1  
US-09-134-001C-5314

Sequence 5314, Application US/09134001C  
Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134, 001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 5314

LENGTH: 930

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5314

Query Match 7.1%; Score 126; DB 2; Length 930;

Best Local Similarity 21.0%; Pred. No. 0.0057;

Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY	9	SDMTGVSTFVILETGLDNCNIYANGLMIG--VIINPTDDEGNFVDIDVT-----	60
DB	436	SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNNISGNGDGGTI-IDSTTIKRVYK	489
QY	61	-----LNDIKIKVDYIDSGIDSGDWFTGNPNENYNT-----IPNS 97	
DB	490	VEGNQMLPDSNRIVDYSEVEDVT-NDYQAQGNNDVNIIPGNDSPYIKVISKIDPNK	548
QY	98	QSYSLKSENSQITQIKRYVSCNTRSLRTKFSARV--TTTSG-----KVISI-- 144	
DB	549	DDYTTIQVTFMQTTINEY-----TGEPRYASVDNTIAFSTSGGCGDLPPEKTYKIGD	603
QY	145	-----TQNSINSSRVVAINAIDATNPDTDELATTKETRENGSYTHKSTNSL-Y 133	
DB	604	YWEDEVDDKQGIQNTNKEPLSNVLTLYTPDG--TSKSVRTDEGKQOPDGLKNGLTY	660
QY	194	VHTTTPRSIKLQNMWEDY-----NNGMTWAQSCYYTTCADGSGEST 236	
DB	661	KITFETP-----EGTTPTKSKSGTNPALDSRGNVWVITINQDDMTIDSGFYQTP	710
QY	237	RWLAQSIRPPGNYDGLMINDIALSGMAHKSYNVDITGINSFTRIICKGFSWYNISG	236
DB	711	KYSIGNVVYDNTKQIGDDEKGISGV--KVLTKDENGNIISTTTDENG--KYQFDN 765	

```
QY      297 LDRGHAVIIDIQ 308
      | : : : |
Db      766 LNSGNVIVHFDK 777

RESULT 2
US-09-386-962C-10
; Sequence 10, Application US/09386962C
; Patent No. 6635473
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLOCOCCUS
; FILE REFERENCE: P06335US2/BAS
; CURRENT APPLICATION NUMBER: US/09/386,962C
; CURRENT FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/098,443
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/117,119
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-386-962C-10

Query Match      7.1%; Score 126; DB 2; Length 930;
Best Local Similarity 21.0%; Pred. No. 0.0057;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY      9 SDMTGVSTFNVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT-----60
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      436 SMFTNIDTKHHTVEQ-----TIYINPLRYSAKETNNVINSNGDEGSTI-IDDSIIIRYK 489
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      61 -----LNDNIXIVYIDSGSDGDMFTGNPNENT-----IPNS 97
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      490 VGDQNLPPDSNRIVDYSEYEDVT-NDDYAQLGNNDVNIINGNIDSPYIIKVISKYPNK 548
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      98 QSYSLLKSENSQITQIRYVSCNTSRLRTKSPSAKV--TTTSG-----KVISI-- 144
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      549 DDTTIOQTYMTTINNEY-----TGEFRFASYNNTIAFTSSGQGGDLPEPETYKIGD 603
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      145 -----TONSINSSRVVINAIDATNFTDDELRTTKETRFENOSYTSKSTNSL-Y 193
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      604 YWVEDVKDGIQNTNNEKPLSNVLVLTYPDG---TSKSVRTDEGKYQFDGKNGLTLY 660
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      194 VHTWTIPRSLKIQNRWEDY-----NNGWTAAQSCYYKTGADGSEST 236
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      661 KITFETP-----EGYTPFLKHSCTNPALDSEGNVWVTINGDDMTIDSGFYQTP 710
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      237 RWLAAGSIFPPGNYDGLMLNDIALSGMAHKSYNVDGINQLSFTRIIKGFVWVNI 296
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      711 KYSIGNTVWYDTNKDGIQSDDEKGISGV--KVTLKDENGNIISTTTDENG---KYQPDN 765
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      297 LDRGHAVIIDIQ 308
      | : : : |
Db      766 LNSGNVIVHFDK 777

RESULT 3
US-09-147-405B-15
; Sequence 15, Application US/09147405B
; Patent No. 6733758
; GENERAL INFORMATION:
; APPLICANT: Guss, Bengt
; APPLICANT: Nilsson, Martin
; APPLICANT: Frykberg, Lars
; APPLICANT: Flock, Jan-Ingmar
; APPLICANT: Lindberg, Martin
; TITLE OF INVENTION: Fibrinogen Binding Protein Originating from
; TITLE OF INVENTION: Coagulase-Negative Staphylococcus
; FILE REFERENCE: guss 09/147405
```

```
      ; CURRENT APPLICATION NUMBER: US/09/147,405B
      ; CURRENT FILING DATE: 1999-04-11
      ; PRIOR APPLICATION NUMBER: PCT/SE97/10191
      ; PRIOR FILING DATE: 1997-06-18
      ; PRIOR APPLICATION NUMBER: SE 9602496-3
      ; PRIOR FILING DATE: 1996-06-20
      ; NUMBER OF SEQ ID NOS: 15
      ; SOFTWARE: PatentIn Ver. 2.0
      ; SEQ ID NO 15
      ; LENGTH: 1092
      ; TYPE: PRT
      ; ORGANISM: Staphylococcus epidermidis
US-09-147-405B-15

Query Match      7.0%; Score 124; DB 2; Length 1092;
Best Local Similarity 21.0%; Pred. No. 0.011;
Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

QY      9 SDMTGVSTFNVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT-----60
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      439 SMFTNIDTKHHTVEQ-----TIYINPLRYSAKETNNVINSNGDEGSTI-IDDSIIIRYK 492
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      61 -----LNDNIXIVYIDSGSDGDMFTGNPNENT-----IPNS 97
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      493 VGDQNLPPDSNRIVDYSEYEDVT-NDDYAQLGNNDVNIINGNIDSPYIIKVISKYPNK 551
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      98 QSYSLLKSENSQITQIRYVSCNTSRLRTKSPSAKV--TTTSG-----KVISI-- 144
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      552 DDTTIOQTYMTTINNEY-----TGEFRFASYNNTIAFTSSGQGGDLPEPETYKIGD 606
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      145 -----TONSINSSRVVINAIDATNFTDDELRTTKETRFENOSYTSKSTNSL-Y 193
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      607 YWVEDVKDGIQNTNNEKPLSNVLVLTYPDG---TSKSVRTDEGKYQFDGKNGLTLY 663
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      194 VHTWTIPRSLKIQNRWEDY-----NNGWTAAQSCYYKTGADGSEST 236
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      664 KITFETP-----EGYTPFLKHSCTNPALDSEGNVWVTINGDDMTIDSGFYQTP 713
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      237 RWLAAGSIFPPGNYDGLMLNDIALSGMAHKSYNVDGINQLSFTRIIKGFVWVNI 296
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      714 KYSIGNTVWYDTNKDGIQSDDEKGISGV--KVTLKDENGNIISTTTDENG---KYQPDN 768
      | : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      297 LDRGHAVIIDIQ 308
      | : : : |
Db      769 LNSGNVIVHFDK 780

RESULT 4
US-09-134-000C-4697
; Sequence 4697, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4697
; LENGTH: 899
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4697

Query Match      6.7%; Score 119; DB 2; Length 899;
Best Local Similarity 19.4%; Pred. No. 0.025;
Matches 74; Conservative 58; Mismatches 119; Indels 130; Gaps 20;

QY      21 LETGLDN-----CNINYANGLMIGVIINITPTDDEGNFVDID----DVTLNDNIXIVD 69
```

Db 144 MKKGLSRVAVATVAVVANDPULPG-----KNTIDVSNMGDISVAEYKIKS 192  
Qy 70 Y-IDGSDIDGSDGFTGNENYNTIPNSOSYSLKSENSQTOIKRYVSCNSTRK 128  
Db 193 YGTGVSVKLTETGVVY-NPYAAGQIRNAKAGIKYS---AYHYSVSA- TAODEAR 246  
Qy 129 SPKAYTTTSGKYSTIGNSIRVYVNAIDATNFDDELRTTKERFENOSYTSKKS 188  
Db 247 YF-AQAAANSQ-----LDKXTIWFDAEDPTLTNNRMAHANSVAFNOOLKALGYK 296  
Qy 189 TNSLYVHTWTI-----PRSLKLNRMWEDYNNGMTAOSCY 225  
Db 297 NDALVYKMWLTNGYIDTSAFGDRVVAQYPTPS---SMQNNHDGAMQMSQMYF 352  
Qy 226 KTGADGSESTRWLAAGSTPPG--NYDGLMNDIALSGMAH-----KSY---- 269  
Db 353 -----PGLANYEGRPDISMTYSNPLMGNSSGPDLSKYTYTP 391  
Qy 270 -----NVDGINGLSF-----TRIGKFSVYVNSIGLDGHAVIDQ--YG 310  
Db 392 GRVIMKNDTFYQDVARTGMRVKKNTLVTKIGIR--YSSAGIPR---LVTDQGYLA 445  
Qy 311 NKXRIIFHAGYENSDEPLSSS 331  
Db 446 NKDYVL--AAGSNIDLYFTTN 464

## RESULT 5

US-10-080-505-17  
; Sequence 17, Application US/10080505  
; Patent No. 6676948  
; GENERAL INFORMATION:  
; APPLICANT: St. Gene, Joseph W.  
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR  
; CURRENT APPLICATION NUMBER: US/10/080,505  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 08/296,791  
; PRIOR FILING DATE: 1994-10-25  
; PRIOR APPLICATION NUMBER: US 09/839,996  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 17  
; LENGTH: 1411  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
; US-10-080-505-17

Query Match 6.5%; Score 115.5; DB 2; Length 1411;  
Best Local Similarity 20.8%; Pred. No. 0.11;  
Matches 85; Conservative 48; Mismatches 131; Indels 145; Gaps 21;  
Qy 4 REDMSDMTGV-----TENVILEGLDNCIYANGLMIGVINIT----- 45  
Db 554 KDLAFNGMFGDKDXTKATGLANTYINPL--NKNHFLSGGTYNKG--NITDGGTIV 607  
Qy 46 ----PTDEGNFVD-----IDVTLLNDNIKIYDY-IDG-----SDIDG 78  
Db 608 FSGRPTTHAVNHLRLNELGRKGEVVIDDMVIRTFKAENFQIKGSGTVVSRVSIIEG 667  
Qy 79 SDGFTYGNPN-EVNTIPNSOSYSLKSENSQTOIKRYVSCNSTRKTSFSAKTTT 137  
Db 668 N-WTISNNANATGAVVPOONTICTRSDWTLT-----TCKTVNLTKKVIDSIPTQ 719  
Qy 138 SGKYSTIONS-----INSGRVIN-----AIDATN- 163  
Db 720 INGSINLNNATVNIHGLAKINGVITLJNSQFTLSNNAOTGNTIQLSNANATVDANL 779  
Qy 164 -----FTDDELRTTKERFENOSYTSKKSSTNSLYVHTWTIPRSLKLNRMWEDYNNGMT 218  
Db 780 NGNVHLTDSAQFSLKNSHFSHQ-IQGXKDTVTLENNATWTPSDTTLQNL---TLNNSIV 835

Qy 219 WAOSCY-----YKTGADGSESTRW---LAAGSIFPGYDGL-----WLDN 257  
Db 836 TNSAYSASNNAPRRRLBETTTPTSEBHFNTLTVNGKLSGCTPFTSSLFQYKSD 895  
Qy 258 DIALSGMAKSYNV---DTG-----INQSLFTRIIGKFSWYVNSIGLD 298  
Db 896 KIKLSNDAGDYTLAVRDTGKEPVTLEQLTL-----IEGLD 931

## RESULT 6

US-09-107-433-5178  
; Sequence 5178, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arinello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5178:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 360 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULAR TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (B) LOCATION 1...360  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5178:  
US-09-107-433-5178

Query Match 6.4%; Score 114; DB 2; Length 360;  
Best Local Similarity 25.6%; Pred. No. 0.019;  
Matches 46; Conservative 33; Mismatches 73; Indels 28; Gaps 8;

Qy 31 YANGLMNIGVILN--TTPDDEGNFVDIDVTLLNDNIKIYDIDG-----SDIDG 78  
Db 104 YKRAHQAISVTLKRGALPIINENDSVVIDEVKGDWDTLSAQVAAVQADLVLLTDVDS 163  
Qy 79 SDGFTYGNENYNTIPNSOSYSLKSENSQTOIKRYVSCN-TSLRTKTSFSAKTTT 137  
Db 164 -----LVYGNP---NSDPRAKRLRIETINREIIDMAGAGASSNGTGMLTKIKAAATATE 216

Qy 138 SGKVISTONSINSRVINAIDATNFTDDE---LRTTKERFENOSYTHSKSTNSLYV 194  
Db 217 SGVPEVYIC-SSLSKD---SWIEBAETEDSGSYFAOEKGLTKOKWLAFAVQSGSISIW 271

RESULT 7  
US-09-583-110-5138

; Sequence 5138, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PAT400-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 5138  
; LENGTH: 376  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-5138

Query Match 6.4%; Score 114; DB 2; Length 376;  
Best Local Similarity 25.6%; Pred. No. 0.02;  
Matches 46; Conservative 33; Mismatches 73; Indels 28; Gaps 8;

Qy 31 YANGLMIGVIIN--ITPTDEGNFVDIDVTLDNLIKIVYIDG-----SDIDG 78  
Db 120 YKNAHQALSVILNKGALPIINENDSVVIDEYKVGNDPLLSAQVAMVQADLLVLTVDG 179  
Qy 79 SDGPFYTGPNPEVNTIPNSQSYSLKSENQITQIKRVSCSN--TSRLRTKSFSAKYTTT 137  
Db 180 ---LYTGNP---NSDPKAKLERIETINREIIDPAGAGSSNGGKWTIKKATYATE 232  
Qy 138 SGKVISTONSINSRVINAIDATNFTDDE---LRTTKERFENOSYTHSKSTNSLYV 194  
Db 233 SGVPEVYIC-SSLSKD---SWIEBAETEDSGSYFAOEKGLTKOKWLAFAVQSGSISIW 287

RESULT 8  
US-10-080-505-13

; Sequence 13, Application US/10080505  
; Patent No. 6676948  
; GENERAL INFORMATION:  
; APPLICANT: St. Gene, Joseph W.  
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
; FILE REFERENCE: A-59941-1/RFA/DCF/DHR  
; CURRENT APPLICATION NUMBER: US/10/080,505  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 08/296,791  
; PRIOR FILING DATE: 1994-10-25  
; PRIOR APPLICATION NUMBER: US 09/839,996  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 1436  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-080-505-13

Query Match 6.3%; Score 112.5; DB 2; Length 1436;  
Best Local Similarity 21.1%; Pred. No. 0.22;  
Matches 64; Conservative 39; Mismatches 103; Indels 97; Gaps 14;

Qy 4 REDSMSPMTGVSYFN-----VILETGDNQNIYANGLMIGVIINIT----- 45  
Db 570 KETLAINVMFPEETDENKNGRLNIKYPTTEDRLILSGGNIWG---NITQEGSTLVFSG 626  
Qy 46 -PTDDEGNFVD-----IDVTINDNIKIVDY-IDG-----SDIDGSDG 81  
Db 627 RPTPHAYNHLNRDNEIGRPOGEVVIDDMITRTFKENFOIKGSAVAVSRVSSIEGN-- 684  
Qy 82 WFTYGNPN-EYNTIPNSQSYSLKSENQITQIKRVSCSNTSRLRTKSFSAKYTTTSGK 140  
Db 685 WTVSNNAAPAGVAPVQONTICTRSDDWTGLT-----TCKTVDLTDTKVINSIPTTING 738  
Qy 141 VISITONS-----INSRVIVIN-----AIDATN----- 163  
Db 739 SIMLTDATYNIHGLAKNGVNLINHSQPTLSNNATQGNQLDSNANATVDANLGN 798  
Qy 164 -FTDDELRTTKERFENOSYTHSKSTNSLYVHTTTPISLKLQNMREDDYNGGWTWAO 221  
Db 799 VHLTDSAQPSLKNSHFSHQ-IOGDKDVTYLENATWTWMPSDATLQNT---TLNNSVTYTLN 854  
Qy 222 SCY 224  
Db 855 SAY 857

RESULT 9  
US-09-830-230A-418

; Sequence 418, Application US/09830230A  
; Patent No. 6902893  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Lyme Disease Vaccine  
; FILE REFERENCE: PB481US  
; CURRENT APPLICATION NUMBER: US/09/830,230A  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: PCT/US98/12718  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/057,483  
; PRIOR FILING DATE: 1997-09-03  
; PRIOR APPLICATION NUMBER: 60/053,344  
; PRIOR FILING DATE: 1997-07-22  
; PRIOR APPLICATION NUMBER: 60/053,377  
; PRIOR FILING DATE: 1997-07-22  
; PRIOR APPLICATION NUMBER: 60/050,359  
; PRIOR FILING DATE: 1997-06-20  
; NUMBER OF SEQ ID NOS: 756  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 418  
; LENGTH: 324  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-830-230A-418

Query Match 6.0%; Score 107.5; DB 2; Length 324;  
Best Local Similarity 18.4%; Pred. No. 0.067;  
Matches 44; Conservative 38; Mismatches 80; Indels 77; Gaps 9;

Qy 14 VSTFNVLFTGLDNCNIYANGLMIGVIINTPTDDEGNFVDIDVTLDNLIKIVDYIDG 73  
Db 122 INYKFTLNGIDILN-----NEKGNFYNALSLIED---VQDY--- 158  
Qy 74 SDIDSGWPFYTGPNPEVNTIP-----NSQSYSLKSENQITQ 112  
Db 159 -----DSYFYF---KFLSIPRAHLKIDSDYFNVTKINYNPPEVYVRNIGDLIQD 210  
Qy 113 IKRVYSCSNTSRL-----RTKSFSAKYTTTSGKVISTONSINSRVINAIDATNFTDD 167  
Db 211 VKNFVLSGNTSKLINIRDKNFFIQSWDKGK-----SNSINTNSFLTWTIRLGGREKN 265  
Qy 168 ELRTTKERFENOSYTHSKSTNSLYVHTW-----TIPSLKLQNMREDDYNGGWTW 219  
Db 266 GIQPAKHLSDSDISYLESRGWDHTHEWFFVKRIYVPRD-----PEINNGWTW 316

```
RESULT 10
US-09-830-230A-417
; Sequence 417, Application US/09830230A
; Patent No. 6902893
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PA481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 417
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-830-230A-417

Query Match      6.0%; Score 107.5; DB 2; Length 343;
Best Local Similarity 18.4%; Pred. No. 0.073;
Matches 44; Conservative 36; Mismatches 80; Indels 77; Gaps 9;

QY 14 VSTPENVLEGLDNCNIYANGLMIGYIINTPPDDEGNFVDIDVTLDNLIKVDYIDG 73
DB 141 INYKFLINNGIDILN-----NEKGNFYFNALSLIED---VQDY--- 177
QY 74 SDIDSGDGFYTGPNENYNTIP-----NSQSYLLKSNISQITQ 112
DB 178 -----DESYFY--KXFLSIPRAHLKIDSDYFNVTCKINYNFNPVVRNGDLID 229
QY 113 IKRYVSCSNTSRL-----RTKSPSAKYTTTSGKVISITONSINSSRVVINAIDATNFTD 167
DB 230 VKQNFVLSGNTSKLININDKNNPFIQSWDQKGGK-----SNSINNSFLTITMIRLGGRRKN 264
QY 168 ELRTKTRFBNOSYTHKSTNSLYVHTW-----TTPRSKLQWNRWEDYNNGW 219
DB 285 GIDFAKHLBADSSDDISYLSERGWMDHIEWYFVKRIYVPRD-----PEINNGWTW 335

RESULT 11
US-09-251-645-5
; Sequence 5, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
```

```
US-09-251-645-5

Query Match      5.9%; Score 106; DB 2; Length 258;
Best Local Similarity 21.3%; Pred. No. 0.066;
Matches 57; Conservative 44; Mismatches 96; Indels 70; Gaps 13;

QY 29 NIYANGLMIGYIINTPTD-DEGNFVD-----IDVTLDNLIKVDYID 72
DB 16 NPYANGRHQCMVATSVLKQYRNGDWIKALSLAEKRSIOVALSBSLIDQLKM----- 70
QY 73 GSDIDSGDGFYTGPNENY-----NTIPNSQSY-----SLKSENSQITQ----- 112
DB 71 -----PSGWTTDANKFPLGLNGVHADAFIDQVTDRAQDCCCTNENYQNSVKSVE 124
QY 113 -IKRYVSCSNTSRLRKSPSAKYT--TSGKVISITONSIN-----SRVYVINAIDATNFT 165
DB 125 IIRYVSSNRTS---TEYLMAKMTFBDQDKRLLTNTMSVGDEVPFSKVLKIAAPYAIN 181
QY 166 DDELRTTKETRFBNOSYTHKSTNS--LYVHTWTIPRSIKLQWNRWEDYNNGWTAQSC 223
DB 182 TNOLEHINITLFPKTEBPT-KSDTHHQIITLVYHTIPYHLRILEGNDSTVNR----- 232
QY 224 YKTTGADGSGESTRWLAAGSIFPPGNY 250
DB 233 IYVLGKEPSND--RFLTRGRVFRKGT 257

RESULT 12
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match      5.9%; Score 105; DB 2; Length 1430;
Best Local Similarity 22.5%; Pred. No. 1.1; Indels 110; Gaps 19;
Matches 82; Conservative 40; Mismatches 133;

QY 5 EDMSDMDTGVSTPENVLEGLDNCN--IYANGLMIGYIIN--ITPTDDEGNF---VD 55
DB 411 EDNSS-----GGYDFILANDIDNSNPVQAELQNLWHLNYSIYANDPEANFDGVRVD 465
QY 56 IDVTLDNLIKIV-DYIDGS-DIDGS-----DGWFTGPNENY-----TI 94
DB 466 AVDNVANADLQIADSVLKAHYGVDSKSEKNAINHLSTLEAW--SDNDPOYKDKYGAQLPI 523
QY 95 PNSQSYSL-----LKSNSQITQIKRYVSCSNTSRLRKSPSAKYTTTSGKVISITON 147
DB 524 DNKRLSLVYALTRPLEKDSNKNRSGLEPYITNSLNRRSABGKNSEMANIYIFRAH 563
QY 148 SINSSRVVINA-----DATNFTDEIRT-----TKETRENOSYT----- 183
DB 584 DSEVQVIAKIITIAQINPKTDGLFTLDELKQAFKLYNEDMRQAKKTYGNSNPTAYALM 643
QY 184 -SHKSTNSLY-----VHTWTIPR-----SLKQWNRWEDY 213
DB 644 LSNKDSITRLLYGDMYSDGQYVATKSPYDAIDITLKKARIKXAAAGQDMKITIYVEGDKS 703
QY 214 NNQWTWA---QSCYVTGA---DGSESTRWLAAGSIFPPGNYDLWL---DNIDALSG 263
```

[illegible]

```

RESULT 13
US-09-210-361-6
, Sequence 6, Application US/09210361
, Patent No. 6284479
, GENERAL INFORMATION:
, APPLICANT: Nichols, Scott E.
, TITLE OF INVENTION: Substitutes for Modified Starches and
, TITLE OF INVENTION: Latexes in Paper Manufacture
, FILE REFERENCE: 03570R
, CURRENT APPLICATION NUMBER: US/09/210,361
, CURRENT FILING DATE: 1998-12-11
, EARLIER APPLICATION NUMBER: 09/007,999
, EARLIER FILING DATE: 1998-01-16
, EARLIER APPLICATION NUMBER: 08/478,704
, EARLIER FILING DATE: 1995-06-07
, EARLIER APPLICATION NUMBER: 09/009,620
, EARLIER FILING DATE: 1998-01-20
, EARLIER APPLICATION NUMBER: 08/485,243
, EARLIER FILING DATE: 1995-06-07
, EARLIER APPLICATION NUMBER: 09/008,112
, EARLIER FILING DATE: 1998-01-16
, EARLIER APPLICATION NUMBER: 08/482,711
, EARLIER FILING DATE: 1995-06-07
, NUMBER OF SEQ ID NOS: 6
, SOFTWARE: FASTSEQ for Windows Version 3.0
, SEQ ID NO: 6
, LENGTH: 1430
, TYPE: PRN
, ORGANISM: streptococcus mutans
US-09-210-361-6

```

Query Match	Similarity	5.9%	Score 105	DB 2	Length 1430
Best Local Match	Similarity	22.5%	Pred. No. 1.1		
Match 82	Conservative	40	Mismatches 133	Indels 110	Gaps 19
Qy	5	EDSM5DMTGVSTFNVILETGLDNCH--IYANLANNIGVITN--ITPDDGNGF----	VD	55	
Db	411	EDNS5-----GGYDELLANDINDSNPPVQAEQOLMILHYLMANGSIVANDPEANFGVGRVD		465	
Qy	56	IDDTVLNINIKV-YVIDGS-DIDDS-----DGFYTGPNEXN-----TI	94		
Db	466	AVDNNVADLLQIASDYLKAHYGVDSSEKNAIHHLSILEAM--SDNDPQNKDTKGAQLP	523		
Qy	95	PNSQSYSL-----LKSNSQITQIKRYVSCSNTSRLLTKSPSAKVTTSGKVISITON	147		
Db	524	DNKLLSLLYALTRPLEKQASKNKEIRSLSEPIVITNSLNNR5AEGNSRMYANFYIFIRAH	583		
Qy	148	SIN5SRVVINAI-----DATNFPDDELRT-----TKETFPENOSTY-----	183		
Db	584	DSEVOVTYIAKKIKKAQINPKTIDGLFTLIDELKQAFKIYMNDDQAKKRYQOSNIPTAYALM	643		
Qy	184	-SHK5STNSLY-----VHTWTIPR-----SLKUNRMREDEY	213		
Db	644	LSNKOSITRLYIGDMXSDGQYMATSPYYDAIDTLTKRIRKKAAGGQDMKITYVEGDS	703		
Qy	214	NNGWTTWA-----QSCYYTKGA-----DGSSESTRWLAAGSIFPPGANYDGLMT--	DNDIALSG	263	
Db	704	HHMDVYTGVLTVSRVGTGANEAATQSS5EATK--TQ5MAVITSNNP5ELAKINDQKVIYVWG	761		
Qy	264	MAHKS	268		
Db	762	AAHKN	766		

```

US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: prt
; ORGANISM: streptococcus mutans
US-09-740-274-6

```

```

Query March 5.9%; Score 105; DB 2; Length 1430;
Best Similarity 22.5%; Pred. No. 1.1;
Match Local 82; Conservative 40; Mismatches 133; Indels 110; Gaps 19;

QY 5 EDMSDWTVGTSFVILLETGLDNCN--TYANGLMIGVIIN--YTPTDEGNF---VD 55
DB EDNS-----GGYDFELLANDIDNENPVQAQLNWLHYLMYGSIVANDPEANFDGVRVD 465
QY 56 IDDTYLANDNIKIY-DYIDGS-DIDGS-----DGMFTYGNPNPEYN-----TI 94
DB 466 AVDVVNAIDLQIADSYLKAHYGVKSEKNAIINHLSILEAW--SNDPQYNDOTYGAQLPI 523
QY 95 PMSOSVSP-----LKSENSQITQIKRYVSCSNTSRLRTKSPFAKYTTSGKYSISITON 147
DB 524 DNKRLSTLYLALTRPLEKDSANKKEIISGLEPVTNSLNNRSABGKNSERMANYPFIABH 583
QY 148 SINSRRVYINAI-----DATNFTDELFAT-----TKETRFERQSTT----- 183
DB 584 DSEVOVTYAKIKQAQINFKTDGLFTFDLQAKAFKIYVEDRQAQKTYQSNIPYALVM 643
QY 184 -SHKSSTNSLY-----VHTWTIPR-----SLKLOWRMEDY 213
DB 644 LSNKDOSTRLLYGDMYSDDGQYMAATKSPYYAIDTLKARKIYAAGSGDMKITTYEGSKS 703
QY 214 NNGWTWA---OSCYKTKGA---DGSSTFEMLAGSIFPGANYDGLWL---DNDIALSG 263
DB 704 HMDMDYTGVLTSVRGTGANEATDQSGSATK--TQGMALVITSNNPSLKLANDKIVYMWG 761
QY 264 MAHKS 268
DB 762 AAHKN 766

```

```

      704 HNDWDTYGLTVSRVOTGANEATDQSEATK--TQSMVITSNNSLSLKNDKVIYNNVG 761
      264 MAHKS 268
      762 AAHKN 766

      RESULT 14

```





THIS PAGE LEFT BLANK

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 30, 2006, 09:43:19 ; Search time 20.8022 Seconds  
(without alignments)  
6708.668 Million cell updates/sec

Title: US-09-889-874a-22

Perfect score: 1784  
Sequence: 1 FTIRSDMSMDWTGVSTFNV.....ILFHAGYNSDPYLSSTIVY 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA Main:  
1: /cgn2\_6/ptodaca/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodaca/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodaca/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodaca/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodaca/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodaca/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126	7.1	892	4 US-10-282-122A-70481	Sequence 70481, A
2	126	7.1	892	4 US-10-661-809-21	Sequence 21, App1
3	126	7.1	930	4 US-10-615-383-10	Sequence 10, App1
4	126	7.1	930	4 US-10-680-184-10	Sequence 10, App1
5	126	7.1	930	4 US-10-689-082-10	Sequence 10, App1
6	126	7.1	930	4 US-10-724-972A-5309	Sequence 5309, App1
7	124	7.0	1092	4 US-10-378-674-8	Sequence 8, App1
8	124	7.0	1092	4 US-10-806-288-15	Sequence 15, App1
9	118.5	6.6	670	4 US-10-282-122A-70444	Sequence 70444, A
10	118	6.6	6761	5 US-10-732-823-15035	Sequence 15035, A
11	115.5	6.5	1411	4 US-10-080-505-17	Sequence 17, App1
12	115.5	6.5	1411	4 US-10-687-046-17	Sequence 17, App1
13	114	6.4	360	5 US-10-617-320-5178	Sequence 5178, App1
14	112.5	6.3	1436	4 US-10-080-505-13	Sequence 13, App1
15	112.5	6.3	1436	4 US-10-687-046-13	Sequence 13, App1
16	111	6.2	369	5 US-10-472-928-1818	Sequence 1818, App1
17	109.5	6.1	285	4 US-10-452-024-65	Sequence 65, App1
18	109	6.1	807	4 US-09-820-843A-108	Sequence 108, App1
19	108	6.1	626	5 US-10-739-930-6661	Sequence 6661, App1
20	108	6.1	978	4 US-10-282-122A-75356	Sequence 75356, A
21	107.5	6.0	324	5 US-10-994-726-418	Sequence 418, App1
22	107.5	6.0	324	5 US-10-994-726-417	Sequence 417, App1
23	107.5	6.0	623	4 US-10-452-024-73	Sequence 73, App1
24	107.5	6.0	667	4 US-10-425-115-292261	Sequence 292261, App1
25	107.5	6.0	753	4 US-10-425-114-54358	Sequence 54358, A
26	107.5	6.0	1038	4 US-10-282-122A-43827	Sequence 43827, A
27	107	6.0	1449	4 US-10-282-122A-65711	Sequence 65711, A

28	106.5	6.0	623	4 US-10-452-024-72	Sequence 72, App1
29	106.5	6.0	718	4 US-10-425-115-332566	Sequence 332566, A
30	106.5	6.0	639	4 US-10-425-114-56258	Sequence 56258, A
31	105.5	5.9	991	5 US-10-831-070-54	Sequence 54, App1
32	105	5.9	507	5 US-10-712-533A-10	Sequence 10, App1
33	105	5.9	1430	3 US-09-740-274-6	Sequence 6, App1
34	105	5.9	1430	3 US-10-383-930-36	Sequence 36, App1
35	105	5.9	1430	5 US-10-797-821-36	Sequence 36, App1
36	104.5	5.9	643	4 US-10-369-493-15821	Sequence 15821, A
37	104.5	5.9	643	4 US-10-369-493-15453	Sequence 15453, A
38	104.5	5.9	699	4 US-10-369-493-16202	Sequence 16202, A
39	103.5	5.8	582	4 US-10-806-288-13	Sequence 13, App1
40	103.5	5.8	593	4 US-10-806-288-11	Sequence 11, App1
41	103.5	5.8	1031	4 US-10-369-493-22004	Sequence 22004, A
42	103.5	5.8	1435	5 US-10-484-703-32	Sequence 32, App1
43	103.5	5.8	1861	5 US-10-472-928-1236	Sequence 1236, App1
44	103	5.8	1395	4 US-10-080-505-7	Sequence 7, App1
45	103	5.8	1395	4 US-10-687-046-7	Sequence 7, App1

## ALIGNMENTS

### RESULT 1

US-10-282-122A-70481  
Sequence 70481, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 70481  
LENGTH: 892  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-10-282-122A-70481

[illegible]

```

RESULT 2
US-10-661-809-21
; Sequence 21, Application US/10661809
; Publication No. US20040101919A1
; GENERAL INFORMATION:
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FROM
; TITLE OF INVENTION: POSITIVE BACTERIA
; FILE REFERENCE: P07741US01/BAS
; CURRENT APPLICATION NUMBER: US/10/661,809
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: 60/410303
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 892
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
US-10-661-809-21

```

Query Match	7.1%;	Score 126;	DB 4;	Length 892;
Best Local Similarity	21.0%;	Pred. No. 0.1;		
Matches	78;	Conservative	51;	Mismatches 141;
				Indels 102;
				Gaps 16;

```

QY 9 GDMGVSFVFNLEELGDNKCNFYAGLMMIG--VIINTPDDEENFADIDVT----- 60
Db SMFTNIDTKNTHVEQ-----TIIINPLKRSYAKETNVNISGNGDEGSTI--IDSTIIKYYK 451
QY 61 -----LNDNIKIYDIDGSDIDSGMFPYTGPNPEYNT-----IPNS 97
Db 452 VGDNGNLDPSNRKIYDYSYEDVT--NDDYAGLGNNNDVNIIFGNIDSPYIIKYSKDEPNK 510
QY 98 QSYLSKSNQIOTQIKRIVSCSNTSRLLTKSFSKAV'-TTTSG-----KVIST-- 144
Db 511 DDTYTIQGVMTQITINEY-----TGEFRTASVDNTIAFSTSGGQGGDLPEPKTYKIGD 565
QY 145 -----TQNSINSRVIVNAIDATNFTDDELRTTYKTFENQSTSHKSSINSL-Y 193
Db 566 YVMEVDVDDKQGLONTNDNEKPLSNVIVTLIYVPGG--TSSXSVTDEGKQVPGDLNGLTY 622

```

CY	194	VHWTTPSPSLKQNMREDEY-----NNGWTAAGSCYYTGTGADGGSSST	236
	:		:
Dd	623	KITFPTP-----EGYFTPLKHSGTNPALDSBGNWSVWTINGDDMTIDBSGFOTP	672
CY	237	RWLAAGSTFFPCPNYDGLMDNDIALSGMAHKSYNVDTGINQLSPTRIIGKFSWVYNISG	296
	:		:
Dd	673	KYSLGNVYWYDINKKGIQGDDKEKGISGV--KYTLKDENGNIISTTTDENG--KYOPDN	727
	:		:
CY	297	LDRGHAVIIDIQ	308
	:		:
Dd	728	INSQNYIVHFDK	739

```

// RESULT 3
// US-10-615-383-10
// Sequence 10, Application US/10615383
// Publication No. US20040038327A1
// GENERAL INFORMATION:
// APPLICANT: POSTER, Timothy
// TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLOCOCCUS
// FILE REFERENCE: P06335US03/BAS
// CURRENT APPLICATION NUMBER: US/10/615,383
// CURRENT FILING DATE: 2003-07-09
// PRIOR APPLICATION NUMBER: 09/386,962
// PRIOR FILING DATE: 1999-08-31
// PRIOR APPLICATION NUMBER: 60/098,443
// PRIOR FILING DATE: 1998-08-31
// PRIOR APPLICATION NUMBER: 60/117,119
// PRIOR FILING DATE: 1999-01-25
// NUMBER OF SEQ ID NOS: 39
// SOFTWARE: Patentin version 3.1
// SEQ ID NO: 10
// LENGTH: 930
// TYPE: PRT
// ORGANISM: Staphylococcus epidermidis
// US-10-615-383-10

```

Query Match 7.1%; Score 126; DB 4; Length 930;  
 Best Local Similarity 21.0%; Pred. No. 0.11;  
 Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

[illegible]

RESULT 4  
US-10-690-184-10

```
Sequence 10, Application US/10690184
Publication No. US20040141997A1
GENERAL INFORMATION:
APPLICANT: POSTER, Timothy
TITLE OF INVENTION: METHODS FOR TREATING OR PREVENTING INFECTIONS FROM COAGULASE-
FILE REFERENCE: P06335US05/BAS
CURRENT APPLICATION NUMBER: US/10/690,184
CURRENT FILING DATE: 2003-10-21
PRIOR APPLICATION NUMBER: 09/386,962
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/098,443
PRIOR FILING DATE: 1998-08-31
PRIOR APPLICATION NUMBER: 60/117,119
PRIOR FILING DATE: 1999-01-25
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 930
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-10-690-184-10

Query Match
Best Local Similarity 21.0%; Score 126; DB 4; Length 930;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGIANNIG--VIINIPPTDEGNFVIDDVT----- 60
DB 436 SMETNIDTGNHVEQ-----TIYNPLRYSAKETNWNISGNDEGSTI-IDDSTIIKVKYK 489
QY 61 -----LNDNIKIVDYIDGSDIDGSDGWFYTGPNENYNT-----IPNS 97
DB 490 VGDNQMLPDSNRIDYSEYEDVT--NDVYAQLGNNNDVNINFGNIDSPYIIKVISKYDPNK 548
QY 98 QSYSLKSENSQITQIKRYVSCNSTRSLRTKFSFAKV--TTTSG-----KVISI-- 144
DB 549 DDYTTIQOTVTWQTINEX-----TGEFRFASVDNTIAFTSSGQGGDLPEPKTYKIGD 603
QY 145 -----TONSINSSRVVINAIDATNFTDDELRTKTRPENOSYSHKSTNSL-Y 193
DB 604 YWEDVDKDGIGQNTNNEKPLSNVLVLTYPDG--TSKSVRTDEGKIQFDGLKXGLTY 660
QY 194 VHTWTIPRSLKLONNWMEDEY-----NNGWTWAOCCYYTGTGADGSEST 236
DB 661 KITFETP-----EGYTPFLKSGTNPALDSEGNVWVTINGQDDMTIDSGFYQTP 710
QY 237 RMLAASIFPPGNYDGLMDNDIALSGMAHKSYNVDGTGINSFTRIIGKGFPMVYNISG 296
DB 711 KYSLGNYVWYDTRKDGIGQDDEKIGISGV--KVLKDENGNIISTTTTDENG--KYQFDN 765
QY 297 LDRGHAIVIIIDQ 308
DB 766 LNSGNYIVHFDK 777

RESULT 5
US-10-689-082-10
Sequence 10, Application US/10689082
Publication No. US20040142348A1
GENERAL INFORMATION:
APPLICANT: POSTER, Timothy
TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLO-
FILE REFERENCE: P06335US04/BAS
CURRENT APPLICATION NUMBER: US/10/689,082
CURRENT FILING DATE: 2003-10-21
PRIOR APPLICATION NUMBER: 09/386,962
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/098,443
PRIOR FILING DATE: 1998-08-31
PRIOR APPLICATION NUMBER: 60/117,119
PRIOR FILING DATE: 1999-01-25
NUMBER OF SEQ ID NOS: 39
```

```
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 930
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-10-689-082-10

Query Match
Best Local Similarity 21.0%; Score 126; DB 4; Length 930;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGIANNIG--VIINIPPTDEGNFVIDDVT----- 60
DB 436 SMETNIDTGNHVEQ-----TIYNPLRYSAKETNWNISGNDEGSTI-IDDSTIIKVKYK 489
QY 61 -----LNDNIKIVDYIDGSDIDGSDGWFYTGPNENYNT-----IPNS 97
DB 490 VGDNQMLPDSNRIDYSEYEDVT--NDVYAQLGNNNDVNINFGNIDSPYIIKVISKYDPNK 548
QY 98 QSYSLKSENSQITQIKRYVSCNSTRSLRTKFSFAKV--TTTSG-----KVISI-- 144
DB 549 DDYTTIQOTVTWQTINEX-----TGEFRFASVDNTIAFTSSGQGGDLPEPKTYKIGD 603
QY 145 -----TONSINSSRVVINAIDATNFTDDELRTKTRPENOSYSHKSTNSL-Y 193
DB 604 YWEDVDKDGIGQNTNNEKPLSNVLVLTYPDG--TSKSVRTDEGKIQFDGLKXGLTY 660
QY 194 VHTWTIPRSLKLONNWMEDEY-----NNGWTWAOCCYYTGTGADGSEST 236
DB 661 KITFETP-----EGYTPFLKSGTNPALDSEGNVWVTINGQDDMTIDSGFYQTP 710
QY 237 RMLAASIFPPGNYDGLMDNDIALSGMAHKSYNVDGTGINSFTRIIGKGFPMVYNISG 296
DB 711 KYSLGNYVWYDTRKDGIGQDDEKIGISGV--KVLKDENGNIISTTTTDENG--KYQFDN 765
QY 297 LDRGHAIVIIIDQ 308
DB 766 LNSGNYIVHFDK 777

RESULT 6
US-10-724-972A-5309
Sequence 5309, Application US/10724972A
Publication No. US20040147734A1
GENERAL INFORMATION:
APPLICANT: Bush, David
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: PAT03-16
CURRENT APPLICATION NUMBER: US/10/724,972A
CURRENT FILING DATE: 2003-12-01
PRIOR APPLICATION NUMBER: 09/450,969
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: 09/134,001
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 7544
SEQ ID NO 5309
LENGTH: 930
TYPE: PRT
ORGANISM: S. epidermidis
US-10-724-972A-5309

Query Match
Best Local Similarity 21.0%; Score 126; DB 4; Length 930;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;

QY 9 SDMTGVSTFNVILETGLDNCNIYANGIANNIG--VIINIPPTDEGNFVIDDVT----- 60
DB 436 SMETNIDTGNHVEQ-----TIYNPLRYSAKETNWNISGNDEGSTI-IDDSTIIKVKYK 489
QY 61 -----LNDNIKIVDYIDGSDIDGSDGWFYTGPNENYNT-----IPNS 97
DB 490 VGDNQMLPDSNRIDYSEYEDVT--NDVYAQLGNNNDVNINFGNIDSPYIIKVISKYDPNK 548
QY 98 QSYSLKSENSQITQIKRYVSCNSTRSLRTKFSFAKV--TTTSG-----KVISI-- 144
DB 549 DDYTTIQOTVTWQTINEX-----TGEFRFASVDNTIAFTSSGQGGDLPEPKTYKIGD 603
QY 145 -----TONSINSSRVVINAIDATNFTDDELRTKTRPENOSYSHKSTNSL-Y 193
DB 604 YWEDVDKDGIGQNTNNEKPLSNVLVLTYPDG--TSKSVRTDEGKIQFDGLKXGLTY 660
QY 194 VHTWTIPRSLKLONNWMEDEY-----NNGWTWAOCCYYTGTGADGSEST 236
DB 661 KITFETP-----EGYTPFLKSGTNPALDSEGNVWVTINGQDDMTIDSGFYQTP 710
QY 237 RMLAASIFPPGNYDGLMDNDIALSGMAHKSYNVDGTGINSFTRIIGKGFPMVYNISG 296
DB 711 KYSLGNYVWYDTRKDGIGQDDEKIGISGV--KVLKDENGNIISTTTTDENG--KYQFDN 765
QY 297 LDRGHAIVIIIDQ 308
DB 766 LNSGNYIVHFDK 777
```

```

Db 436 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNVNISGNDEGSTI-IDDSTIIKVKY 489
QY 61 -----LNDIKIKVDYIDSGDIDSGDMFYTGPNPENT-----IPNS 97
Db 490 VGDNONLPDSNRKYDYSEYEDVT--NDYAOQLGNNDVNINFGNIDSPYIIKVISKDPNK 548
QY 98 QSYSLKSENSQITQIKRYVSCSNTSRLRTKSPSAKV--TTTSG-----KVISI-- 144
Db 549 DDYTTIQOTVMTQTTINEY-----TGEFRITASYDNTIAFSTSSGQGGDLPEPKTYKIGD 603
QY 145 -----TONSINSRVRVINAIDATNFTDDELATTKETREPNOSYTSHKSTNSL-Y 193
Db 604 YWEDVDKDGIOQNTNDEKPLSNVLTLYTPDG---TSKSVRTDEBCKYQFDGLKNGLTY 660
QY 194 VHTWTIPRSLKLGQWREDEY-----NNGMTWAQSCYYTKGADGSEST 236
Db 661 KITFETP-----EGTTPFLKHSCTNPALDSEGNISVWVTINGDDMTIDSGFYQTP 710
QY 237 RWLAAGSIFPPGNYVDGLMNDIALSGMAHKSYNVDTGINQLSFTRIIGKFSWVYNISG 296
Db 711 KYSLGNYVWYDTNKDGIQGDDEKIGISGV--KYTLKDENGNIISTTTTDENG---KYQFDN 765
QY 297 LDRGHAVIITIDQ 308
Db 766 LNSGNYIVHFDK 777

```

## RESULT 7

```

US-10-378-674-8
; Sequence 8, Application US/10378674
; Publication No. US20040006209A1
; GENERAL INFORMATION:
; APPLICANT: PATTI, Joseph M.
; TITLE OF INVENTION: MONOCLONAL AND POLYCLONAL ANTIBODIES RECOGNIZING COAGULASE-NEGATI
; TITLE OF INVENTION: STAPHYLOCOCCAL PROTEINS
; FILE REFERENCE: P07556U01/BAS
; CURRENT APPLICATION NUMBER: US/10/378,674
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: 60/361,324
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1092
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-378-674-8

```

Query Match 7.0%; Score 124; DB 4; Length 1092;

Best Local Similarity 21.0%; Pred. No. 0.2;

Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

```

QY 9 SDMTGVSTRVILLETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT----- 60
Db 439 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNVNISGNDEGSTI-IDDSTIIKVKY 492
QY 61 -----LNDIKIKVDYIDSGDIDSGDMFYTGPNPENT-----IPNS 97
Db 493 VGDNONLPDSNRKYDYSEYEDVT--NDYAOQLGNNDVNINFGNIDSPYIIKVISKDPNK 551
QY 98 QSYSLKSENSQITQIKRYVSCSNTSRLRTKSPSAKV--TTTSG-----KVISI-- 144
Db 552 DDYTTIQOTVMTQTTINEY-----TGEFRITASYDNTIAFSTSSGQGGDLPEPKTYKIGD 606
QY 145 -----TONSINSRVRVINAIDATNFTDDELATTKETREPNOSYTSHKSTNSL-Y 193
Db 607 YWEDVDKDGIOQNTNDEKPLSNVLTLYTPDG---TSKSVRTDEBCKYQFDGLKNGLTY 663
QY 194 VHTWTIPRSLKLGQWREDEY-----NNGMTWAQSCYYTKGADGSEST 236
Db 664 KITFETP-----EGTTPFLKHSCTNPALDSEGNISVWVTINGDDMTIDSGFYQTP 713
QY 237 RWLAAGSIFPPGNYVDGLMNDIALSGMAHKSYNVDTGINQLSFTRIIGKFSWVYNISG 296

```

```

Db 714 KYSLGNYVWYDTNKDGIQGDDEKIGISGV--KYTLKDENGNIISTTTTDENG---KYQFDN 768
QY 297 LDRGHAVIITIDQ 308
Db 769 LNSGNYIVHFDK 780

```

## RESULT 8

US-10-806-288-15

; Sequence 15, Application US/10806288

; Publication No. US20040209326A1

; GENERAL INFORMATION:

; APPLICANT: Guss, Bengt

; APPLICANT: Nilsson, Martin

; APPLICANT: Frykberg, Lars

; APPLICANT: Flock, Jan-Ingemar

; APPLICANT: Lindberg, Martin

; TITLE OF INVENTION: Fibrinogen Binding Protein Originating from

; FILE REFERENCE: Guss 09/147405

; CURRENT APPLICATION NUMBER: US/10/806,288

; CURRENT FILING DATE: 2004-03-23

; PRIOR APPLICATION NUMBER: US/09/147,405

; PRIOR FILING DATE: 1999-04-11

; PRIOR APPLICATION NUMBER: PCT/SE97/10191

; PRIOR FILING DATE: 1997-06-18

; PRIOR APPLICATION NUMBER: SE 9602496-3

; PRIOR FILING DATE: 1996-06-20

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 15

; LENGTH: 1092

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-10-806-288-15

Query Match 7.0%; Score 124; DB 4; Length 1092;

Best Local Similarity 21.0%; Pred. No. 0.2;

Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

```

QY 9 SDMTGVSTRVILLETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT----- 60
Db 439 SMFTNIDTKNHTVEQ-----TIYINPLRYSAKETNVNISGNDEGSTI-IDDSTIIKVKY 492
QY 61 -----LNDIKIKVDYIDSGDIDSGDMFYTGPNPENT-----IPNS 97
Db 493 VGDNONLPDSNRKYDYSEYEDVT--NDYAOQLGNNDVNINFGNIDSPYIIKVISKDPNK 551
QY 98 QSYSLKSENSQITQIKRYVSCSNTSRLRTKSPSAKV--TTTSG-----KVISI-- 144
Db 552 DDYTTIQOTVMTQTTINEY-----TGEFRITASYDNTIAFSTSSGQGGDLPEPKTYKIGD 606
QY 145 -----TONSINSRVRVINAIDATNFTDDELATTKETREPNOSYTSHKSTNSL-Y 193
Db 607 YWEDVDKDGIOQNTNDEKPLSNVLTLYTPDG---TSKSVRTDEBCKYQFDGLKNGLTY 663
QY 194 VHTWTIPRSLKLGQWREDEY-----NNGMTWAQSCYYTKGADGSEST 236
Db 664 KITFETP-----EGTTPFLKHSCTNPALDSEGNISVWVTINGDDMTIDSGFYQTP 713
QY 237 RWLAAGSIFPPGNYVDGLMNDIALSGMAHKSYNVDTGINQLSFTRIIGKFSWVYNISG 296
Db 714 KYSLGNYVWYDTNKDGIQGDDEKIGISGV--KYTLKDENGNIISTTTTDENG---KYQFDN 768
QY 297 LDRGHAVIITIDQ 308
Db 769 LNSGNYIVHFDK 780

```

## RESULT 9

US-10-282-122A-70444

; Sequence 70444, Application US/10282122A

```

Db      282 DTMKDQIGQDDEKGISGV--KVTLLKQNGNIIISTTTDENG---KYQFDNLNSGNYLVHF 338
QY      307 DQ 308
        ||
Db      337 DK 338

RESULT 10
US-10-732-923-15035
; Sequence 15035, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15 (52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 15035
; LENGTH: 6761
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-15035

Query Match      6.6%; Score 118; DB 5; Length 6761;
Best Local Similarity 21.9%; Pred. No. 7.4;
Matches 80; Conservative 49; Mismatches 123; Indels 114; Gaps 18;

QY      17 FNVILETGLDNCNIYANGLMIGVIINIPTDDEGNFVDIDV--TLNDNIKIYDIYDGS 74
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1113 FNCIIIESFENYVDINVEG-----EDGNFYVFKNKLNKIRKQVITIKNETDSS 1158

QY      75 DI-----DGSQGWFTGPNRBNYNTIPNSQSYSLKSENSQITQIK---HYVSGSNTSRL 125
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1159 DMTIELKDBEKGFKTIGYRVSF-----LKKONKKGAQVKGVIKRTYS-----P 1202

QY      126 RTKSPF-----SAKVTTTSGKVISITQNSINSSRVVINAIDATNFTDDELTATTKETRF-- 177
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1203 KYQGFPIHMLENLYKLVTSD--LSKGNNNNNNNM-----KENGFVK 1242

QY      178 ---ENOSYTHKSTNSLYVHTW--TIPRSILKLOWRMEDYDNGWTMAQSCYYTKGADGGS 233
        :||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db      1243 CGHNSYSRACKSKINSISYINKKYVYELNKEITIIIDDDNNNNIN--SCCTYKNNLSNEN 1300

QY      234 ESTRWLAAGSIFPPGNVDGLMLNDIALSGMAHKSVPYVDG---INQISTRIIGKGFSW 290
        :||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db      1301 N---LCADKVLCDGNYN--MLENDV-----DEMNVDONQKKRNDLVSDVKGWSPY 1348

QY      291 VYNIISGIDR-----GHAIVIIDQYGN-----KYRIIFHAGYNSDP 326
        |||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db      1349 FYNIKILNMYKEBERPPILEKCKDKGEMTDHVANNLTNKLKEICISCRILLFMKGHDHYSC 1408

QY      327 YLSSSI 332
        |||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db      1409 SLSSAI 1414

RESULT 11
US-10-080-505-17
; Sequence 17, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RET/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/226,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20

```

NUMBER OF SEQ ID NOS: 58  
 SOFTWARE: Patentin version 3.1  
 SEQ ID NO: 17  
 LENGTH: 1411  
 TYPE: PRT  
 ORGANISM: Haemophilus influenzae  
 US-10-080-505-17

Query Match 6.5%; Score 115.5; DB 4; Length 1411;  
 Best Local Similarity 20.8%; Pred. No. 1.5;  
 Matches 85; Conservative 48; Mismatches 131; Indels 145; Gaps 21;

QY 4 REDSDMTGVS-----TFNVILETGLDNCNIYANGLMIGIYINIT-----45  
 DB 554 RDIAPNGWFGDKDTKMTGRNLAVTYNPL--NKNHFLISGDTLKG---NITDGGTIV 607  
 QY 46 ---PTDDEGNFVD-----IDVTLNDNIKIYDY-IDG-----SDIDG 78  
 DB 608 FSGRPTPHAVNHLNRLNELGRPKGEVVIDDDWIRTFKAENFQIKGGSTVSRNVSSIEG 667  
 QY 79 SDGWFTYGNPV-EYNTIPNSQSYSLKSENSQITQIKRYVSCNMTSLRTKSPSAKYTTT 137  
 DB 668 N-WTISNNANATGVPVNOQNTICTRSDWTGLT-----TCKTVNLTDKKVIDSIPTTQ 719  
 QY 138 SGKVISITONS-----INSSRVIN-----AIDATN- 163  
 DB 720 INGSINLTNATVINHGLAKNGVTLINHSQFTLSNNATQTNIGLSNHNANATVDANL 779  
 QY 164 ----FTDDELRRTKTRFENQSYTSKSSNSTSLYHTWTIPSLKIQNRWEDYNNGWT 218  
 DB 780 NGNVHLTDSAQFSLKNSHFHQ-IOGDKDTVTLENAWTMPSDTTLQNL--TLNNSTV 835  
 QY 219 WAOSCY-----YKTGADGSESTRW---LAASIRPPGNYDGL-----WLDN 257  
 DB 836 TLNSAVSASSNNAPRRHRSLETTPTSEHRFNTLVNKGLSGGQTFQFTSSLSFGYKSD 895  
 QY 258 DIALSGMAHKSYNV---DTG-----INQLSFTRLIGKFSWVYINISGLD 298  
 DB 896 KIKLSNDAEGDYTLAVRDTGKEPVTLEQLTL-----IEGLD 931

RESULT 12  
 US-10-687-046-17  
 Sequence 17, Application US/10687046  
 Publication No. US20040157241A1  
 GENERAL INFORMATION:  
 APPLICANT: St. Geme, Joseph W.  
 TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
 FILE REFERENCE: A-59941-1/RT/DCF/DRR  
 CURRENT APPLICATION NUMBER: US/10/687,046  
 CURRENT FILING DATE: 2003-10-15  
 PRIOR APPLICATION NUMBER: US/10/080,505  
 PRIOR FILING DATE: 2002-02-22  
 PRIOR APPLICATION NUMBER: US 08/296,791  
 PRIOR FILING DATE: 1994-10-25  
 PRIOR APPLICATION NUMBER: US 09/839,996  
 PRIOR FILING DATE: 2001-04-20  
 NUMBER OF SEQ ID NOS: 58  
 SOFTWARE: Patentin version 3.1  
 SEQ ID NO: 17  
 LENGTH: 1411  
 TYPE: PRT  
 ORGANISM: Haemophilus influenzae  
 US-10-687-046-17

Query Match 6.5%; Score 115.5; DB 4; Length 1411;  
 Best Local Similarity 20.8%; Pred. No. 1.5;  
 Matches 85; Conservative 48; Mismatches 131; Indels 145; Gaps 21;

QY 4 REDSDMTGVS-----TFNVILETGLDNCNIYANGLMIGIYINIT-----45  
 DB 554 RDIAPNGWFGDKDTKMTGRNLAVTYNPL--NKNHFLISGDTLKG---NITDGGTIV 607

QY 46 ---PTDDEGNFVD-----IDVTLNDNIKIYDY-IDG-----SDIDG 78  
 DB 608 FSGRPTPHAVNHLNRLNELGRPKGEVVIDDDWIRTFKAENFQIKGGSTVSRNVSSIEG 667  
 QY 79 SDGWFTYGNPV-EYNTIPNSQSYSLKSENSQITQIKRYVSCNMTSLRTKSPSAKYTTT 137  
 DB 668 N-WTISNNANATGVPVNOQNTICTRSDWTGLT-----TCKTVNLTDKKVIDSIPTTQ 719  
 QY 138 SGKVISITONS-----INSSRVIN-----AIDATN- 163  
 DB 720 INGSINLTNATVINHGLAKNGVTLINHSQFTLSNNATQTNIGLSNHNANATVDANL 779  
 QY 164 ----FTDDELRRTKTRFENQSYTSKSSNSTSLYHTWTIPSLKIQNRWEDYNNGWT 218  
 DB 780 NGNVHLTDSAQFSLKNSHFHQ-IOGDKDTVTLENAWTMPSDTTLQNL--TLNNSTV 835  
 QY 219 WAOSCY-----YKTGADGSESTRW---LAASIRPPGNYDGL-----WLDN 257  
 DB 836 TLNSAVSASSNNAPRRHRSLETTPTSEHRFNTLVNKGLSGGQTFQFTSSLSFGYKSD 895  
 QY 258 DIALSGMAHKSYNV---DTG-----INQLSFTRLIGKFSWVYINISGLD 298  
 DB 896 KIKLSNDAEGDYTLAVRDTGKEPVTLEQLTL-----IEGLD 931

RESULT 13  
 US-10-617-320-5178  
 Sequence 5178, Application US/10617320  
 Publication No. US20050136404A1  
 GENERAL INFORMATION:  
 APPLICANT: Lynn A Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
 SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
 THERAPEUTICS

NUMBER OF SEQUENCES: 5206  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: <Unknown>  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: <Unknown>  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/617,320  
 FILING DATE: 10-Jul-2003  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107,433  
 FILING DATE: 30-Jun-1998  
 APPLICATION NUMBER: 60/085131  
 FILING DATE: May 12, 1998  
 APPLICATION NUMBER: 60/051553  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ariniello, Pamela Deneke  
 REGISTRATION NUMBER: 40,489  
 REFERENCE/DOCKET NUMBER: GTC-011  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277  
 INFORMATION FOR SEQ ID NO: 5178:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 360 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Streptococcus pneumoniae



```
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...360
SEQUENCE DESCRIPTION: SEQ ID NO: 5178
US-10-617-320-5178

Query Match
Best Local Similarity 25.6%; Pred. No. 0.33; Length 360;
Matches 46; Conservative 33; Mismatches 73; Indels 28; Gaps 8;

QY 31 YANGLMNIGVYIN--ITPTDEGNFVDIDVTINDNIIKIVYIDG-----SDIDG 78
DB 104 YKNAHQALSTLIRNGALPIINENDSVVIDEVKVDNDPTLSAQVAMVQADLLVLTVDVG 163
QY 79 SDGMFYGNNEVYITINSOSYSLSKSENSQITQIKRYVCSN-TSLRTKSPSAKYTTT 137
DB 164 ----LYTGNP--NSDPRARLERIETINREIIDMAGAGSSNGTGMLTKIKAKATTATE 216
QY 138 SGKVISITONSINSSRVVINAIDATNFTDDE--LRTTKETRFENQSYTHKSSTNSLYV 194
DB 217 SGVPVYIC-SSLSKD---SMIEAERTEDGSYVVAQEKLRTOQKMLAFYAQSGSISIW 271

RESULT 14
US-10-080-505-13
; Sequence 13; Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/236,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 13
; LENGTH: 1436
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-13

Query Match
Best Local Similarity 21.1%; Pred. No. 2.8; Length 1436;
Matches 64; Conservative 39; Mismatches 103; Indels 97; Gaps 14;

QY 4 REDSMDSMTGVSTFN-----VILETGDNINIVANGLMIGVINIT----- 45
DB 570 KEIAYNGWFETIDENKNGRLNIYKPTTDRILLSGGTNLKG--NITQEGGLVFSG 626
QY 46 -PTDDEGNFVD-----IDVTINDNIIKIVY-IDG-----SDIDGSDG 81
DB 627 RPTPHAYNHLNRBNELGRPGQEVVIDDWTTRTFKAFNFOIKGSAVSHVSIEN-- 684
QY 82 WFTYGNPN-EYNTIPNSQSYLSKSENSQITQIKRYVCSNTSLRTKSPSAKYTTTSGK 140
DB 685 WTVSNNAANAAGVVPNOQNTICTRSDWTGLT-----TCKTVDLTDKVINSLIPTQING 738
QY 141 VISITONS-----INSSRVIN-----AIDATN----- 163
DB 739 SIMLIDNATVNIHGLAKNGVTLINHSOPTLSNNATQGNIGLSNNAATVDNANLNGN 798
QY 164 --FTDDELRTTKETRFENQSYTHKSSTNSLYVHTWTIPRSLKIQNWRMEDYNNGWTAQ 221
DB 799 VHLTDSAQFSLKNSHFSHQ-IQGDKDTVTYLENATWTMPSDATLQNL---TLNNSTVTLN 854
QY 222 SCY 224
DB 855 SAY 857
```

```
RESULT 15
US-10-687-046-13
; Sequence 13; Application US/10687046
; Publication No. US20040157241A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/687,046
; PRIOR FILING DATE: 2003-10-15
; PRIOR APPLICATION NUMBER: US/10/080,505
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/236,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 13
; LENGTH: 1436
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-687-046-13

Query Match
Best Local Similarity 21.1%; Pred. No. 2.8; Length 1436;
Matches 64; Conservative 39; Mismatches 103; Indels 97; Gaps 14;

QY 4 REDSMDSMTGVSTFN-----VILETGDNINIVANGLMIGVINIT----- 45
DB 570 KEIAYNGWFETIDENKNGRLNIYKPTTDRILLSGGTNLKG--NITQEGGLVFSG 626
QY 46 -PTDDEGNFVD-----IDVTINDNIIKIVY-IDG-----SDIDGSDG 81
DB 627 RPTPHAYNHLNRBNELGRPGQEVVIDDWTTRTFKAFNFOIKGSAVSHVSIEN-- 684
QY 82 WFTYGNPN-EYNTIPNSQSYLSKSENSQITQIKRYVCSNTSLRTKSPSAKYTTTSGK 140
DB 685 WTVSNNAANAAGVVPNOQNTICTRSDWTGLT-----TCKTVDLTDKVINSLIPTQING 738
QY 141 VISITONS-----INSSRVIN-----AIDATN----- 163
DB 739 SIMLIDNATVNIHGLAKNGVTLINHSOPTLSNNATQGNIGLSNNAATVDNANLNGN 798
QY 164 --FTDDELRTTKETRFENQSYTHKSSTNSLYVHTWTIPRSLKIQNWRMEDYNNGWTAQ 221
DB 799 VHLTDSAQFSLKNSHFSHQ-IQGDKDTVTYLENATWTMPSDATLQNL---TLNNSTVTLN 854
QY 222 SCY 224
DB 855 SAY 857
```

Search completed: January 30, 2006, 09:52:52  
Job time : 22.8022 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2006, 09:46:19 ; Search time 2.66268 Seconds  
(without alignments)  
1358.372 Million cell updates/sec

Title: US-09-889-874a-22

Perfect score: 1784  
Sequence: 1 FTLRDSMSDWTGVSTFNVN.....ILPHAGVNSDPYLSSTIVY 334

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 75621 seqs, 10829074 residues

Total number of hits satisfying chosen parameters: 75621

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

Published Applications AA New:  
1: /cgn2\_6/ptcdatc/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptcdatc/1/pubpaa/US06\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptcdatc/1/pubpaa/US07\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptcdatc/1/pubpaa/PTC\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptcdatc/1/pubpaa/US09\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptcdatc/1/pubpaa/US10\_NEW\_PUB pep.\*  
7: /cgn2\_6/ptcdatc/1/pubpaa/US11\_NEW\_PUB pep.\*  
8: /cgn2\_6/ptcdatc/1/pubpaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	6.6	773	US-11-052-554A-341	Sequence 341, App
2	107	6.0	1449	US-11-052-554A-237	Sequence 237, App
3	101	5.7	3194	US-11-052-554A-90	Sequence 90, App1
4	100.5	5.6	331	US-11-016-564-3	Sequence 3, App1
5	100.5	5.6	829	US-10-880-144-6	Sequence 6, App1
6	99.5	5.6	2399	US-11-052-554A-92	Sequence 92, App1
7	99.5	5.6	2902	US-11-052-554A-91	Sequence 91, App1
8	96.5	5.4	744	US-11-052-554A-94	Sequence 94, App1
9	95.5	5.4	807	US-11-052-554A-55	Sequence 55, App1
10	95.5	5.4	1468	US-10-467-657-1088	Sequence 1088, App
11	95	5.3	980	US-11-052-554A-17	Sequence 17, App1
12	94	5.3	2314	US-11-013-759-11	Sequence 11, App1
13	93.5	5.2	486	US-10-416-047-14	Sequence 14, App1
14	93.5	5.2	710	US-11-089-551A-33	Sequence 33, App1
15	92.5	5.2	1579	US-11-052-554A-9	Sequence 9, App1
16	92.5	5.2	2053	US-11-013-759-9	Sequence 9, App1
17	91.5	5.1	453	US-11-052-554A-224	Sequence 224, App
18	91.5	5.1	539	US-11-069-642-16	Sequence 16, App1
19	91.5	5.1	795	US-11-052-554A-87	Sequence 87, App
20	91.5	5.1	826	US-10-793-626-1066	Sequence 1066, App
21	91	5.1	2340	US-11-052-554A-171	Sequence 171, App
22	90	5.0	395	US-10-793-626-668	Sequence 668, App
23	90	5.0	1250	US-11-052-554A-16	Sequence 16, App1
24	89	5.0	2715	US-11-113-424-51	Sequence 51, App1
25	88.5	5.0	566	US-11-033-039-1244	Sequence 1244, App

26	88.5	5.0	627	US-10-873-528-191	Sequence 191, App
27	88	4.9	1461	US-11-052-554A-283	Sequence 283, App
28	87	4.9	523	US-10-641-678-45	Sequence 45, App1
29	87	4.9	952	US-11-207-626A-45	Sequence 45, App1
30	86.5	4.8	518	US-11-052-554A-355	Sequence 355, App
31	86.5	4.8	1770	US-11-103-957-21	Sequence 21, App1
32	86	4.8	477	US-11-089-551A-34	Sequence 34, App1
33	86	4.8	930	US-10-517-939-106	Sequence 106, App
34	85.5	4.8	878	US-11-052-554A-23	Sequence 23, App
35	85.5	4.8	1067	US-11/062	Sequence 3, App1
36	85.5	4.8	1092	US-11/062	Sequence 6, App1
37	85.5	4.8	1694	US-11-052-554A-83	Sequence 83, App1
38	85.5	4.8	3623	US-10-995-561-593	Sequence 593, App
39	85	4.8	1094	US-10-517-939-144	Sequence 144, App
40	85	4.8	2376	US-11-036-051-4	Sequence 4, App1
41	85	4.8	2715	US-11-036-051-2	Sequence 2, App1
42	84.5	4.7	338	US-11-016-564-11	Sequence 11, App1
43	84	4.7	295	US-10-467-657-5222	Sequence 5222, App
44	84	4.7	857	US-11-052-554A-218	Sequence 218, App
45	83.5	4.7	587	US-11-052-554A-118	Sequence 118, App

## ALIGNMENTS

RESULT 1  
US-11-052-554A-341  
; Sequence 341, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 341  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Shigella flexneri 2a str. 2457T  
; US-11-052-554A-341

Query: Match 6.6%; Score 118; DB 7; Length 773;  
Best Local Similarity 22.7%; Pred. No. 0.032;  
Matches 83; Conservative 55; Mismatches 151; Indels 76; Gaps 18;  
QY 2 TLNRDSMSDWTGVSTFNVNILET-----GLDNCNIYANGIMIGVITNPTDDEG 51  
DB 76 SLDBESTPAD-TGSNNENNAIAKMDMAGEITTHGHSYAAVAN-----GTVVKAGDTLDYT 129  
QY 52 N----FVDIDDTVLNDINIKVYDIIDGSDIDSGMFTTGNPNR-----YN----TINSQ 98  
DB 130 NASVTLDVDTITGDAHAIAAROGT-VFENGEBIYTTGPDAAIAKIYNGVTVTLNXTS 188  
QY 99 SYS-----LASENSQIQIKRYVSCNSRT-----KSPFAKTTTSGKVIYSI 144  
DB 189 AVAHQSGIYLESSINGQ---EATYDILSGSLSRSLANEILYHKDETNSVITIDSEVSSA 244  
QY 145 TONSINSSRVVINAIDATNFTDDELRTTKETRENOYSYTHKSSTNSLYVHTWTIPRSLK 204  
DB 245 ADVFINNIKHL-TVDATN---SKTGSANISITDDNHTYLSIDNS---TMDIKADST 296  
QY 205 LQWRMEDVYNGMTWMAQSCYTKGADGGS-ESTRWLAAGSIFPPGNYDGLMDNDIALSG 263  
DB 297 VSN-----LTVNSTVYISRADGRDVEPRITLTENYV--GNNGVLAHRLTELDNDN 345  
QY 264 MAKSYVVDVINGINQLSFTRITIGKGFVWVNIISGLDGHAVIITIDYGN-----KRIIDFHA 319

Db 346 SAIDKVINNGTSTTVKVTNAGSGATYTLNGIE-----ITSVGENSGEPIIDSRIFAG 401  
QY 320 GYENS 324  
Db 402 AYEYS 406

RESULT 2  
US-11-052-554A-237  
; Sequence 237, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 237  
; LENGTH: 1449  
; TYPE: PRT  
; ORGANISM: *Neisseria meningitidis* Z2491  
US-11-052-554A-237

Query Match 6.0%; Score 107; DB 7; Length 1449;  
Best local similarity 21.1%; Pred. No. 0.54; Mismatches 126; Indels 94; Gaps 17;  
Matches 71; Conservative 46;  
QY 9 SDMTGVSTFNYLRTGLDNCINIVANGIMGVIINITPTDDEGNFVDIDD-----V 59  
Db 722 SDMTGLT--NCEVETITDD-----KVIALSTFTDISGN-VSLADHAHLNLGLA 767  
QY 60 TLNDNIKIVDYIDSGDDIDSGCMFYTGPNENYNTIPN-SQSYSLKSENSQITQIKYVS 118  
Db 768 TLNGLS-----ANGDTRYTVSHANTONGNLSLVGNAQAT-FNQATL 808  
QY 119 CSNSTRLRKTSFSAKVTTTSGKVISITONS-INSRVVINAIDATNPTDDELRTTKETRF 177  
Db 809 NGNTSASGNASFNLSNNAONGSLTSDNAKANVSHSLNCG-----NVSILAKAVFHF 861  
QY 178 ENOSYTSKHSSTNSLYVA-----TWTPRSCLKQWRMEDYNNGWMTAQCYY--KTGADG 231  
Db 862 ENSRFTQGLSGSKDTAHLKXOSEWTLPGCTELGNL--NLDNATITLNSAVRHDAAGQT 918  
QY 232 GS-----ESTRWLAAGSITPPGNYDGLMDNDIALSGMAHKSYNVDGTGINSFRII 284  
Db 919 GSVSDTPRRRRSRRL--SVTPPTSVESRF--NTLTVNGK-----LNGGTFRFM 964  
QY 285 GKGSFV-----VYNISGLDGHAVIITDQ 308  
Db 965 SELFGYRSDKLAESSSEGYTTLAVNNTGNEPVSIDQ 1001

RESULT 3  
US-11-052-554A-90  
; Sequence 90, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 90  
; LENGTH: 3194  
; TYPE: PRT  
; ORGANISM: *Helicobacter pylori* J99  
US-11-052-554A-90

Query Match 5.7%; Score 101; DB 7; Length 3194;  
Best local similarity 18.6%; Pred. No. 4.4;  
Matches 84; Conservative 51; Mismatches 140; Indels 176; Gaps 22;  
QY 12 TGVTSTNVILETGLDNCIYA--NG-----LNMIGVI-INITPTDDEGNFVDIDVTL 61  
Db 561 TGFTFADRYVTG---NMWSGNGAQTGGGATNPGVATIEINIA---GATPKLKTTSQ 612  
QY 62 ND-----NIKIVDYIDSGDDIDSGCMFYTG----- 87  
Db 613 NSYMTFMALONGSGSGKINVSQSDFYDWD---GGYDFGNGVPSVNFNKAYYKFGA 668  
QY 88 PNEYN-----TIPNSQSYSLKSENSQITQIK-----RYVSCS 120  
Db 669 ENSYNEFKNTFLAGNFKQKTIIEKSVLNDASYA--FDGVNNAFNEDEKFGSGSEFNHAB 727  
QY 121 NTSRLRKTSPSAVTTTSGKVISITONSINSRVVINAIDATNPTDDELRTTKETRF-- 178  
Db 728 QTNAFNNNSPSSGSPFSAKQVDPNGNSFNGVAFNFNTPKASFTINDTFVNNQFKLNGA 787  
QY 179 -----NOSYTSKHSSTNSLY-VHTW----- 197  
Db 788 QTFPTSKGVFNMQGLSLSVGTTYQLNMAKSVGKDNMNLQMLRWTSGENPSGKL 847  
QY 198 -----TTPRSCLKQWRMEDYNNGWMTAQCYYTKTGADGSESRWLAAQSITPPGNYDG 252  
Db 848 VDEKKTAPNSAKIYNQFTD--NGLTY---YIKENFNNGIITLRLCTLTGYHCN--- 897  
QY 253 LMLDNDIALSGMAHKSYNVDGTGINSFTRIIGKFSFVYNNISGL-----DRGHAVIITDQ 308  
Db 898 --IDND-----ANLKNVNNNNSNT-----VYLYNGMTTWKTRAGTGVFPDY 937  
QY 309 YGNKYRILFHAGYENSDDPYL-----SSSIV 333  
Db 938 SGTNSVLVFN---QTFPLAGANPTNSNVV 964

RESULT 4  
US-11-016-564-3  
; Sequence 3, Application US/11016564  
; Publication No. US20050287146A1  
; GENERAL INFORMATION:  
; APPLICANT: INHIBITEX, INC.  
; TITLE OF INVENTION: METHOD OF INHIBITING CANDIDA-RELATED INFECTIONS  
; FILE REFERENCE: P08140US03/BAS  
; CURRENT APPLICATION NUMBER: US/11/016,564  
; CURRENT FILING DATE: 2004-12-20  
; PRIOR APPLICATION NUMBER: US 60/566,082  
; PRIOR FILING DATE: 2004-04-29  
; PRIOR APPLICATION NUMBER: US 60/561,540  
; PRIOR FILING DATE: 2004-04-13  
; PRIOR APPLICATION NUMBER: US 60/530,654  
; PRIOR FILING DATE: 2003-12-19  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 3  
; LENGTH: 331  
; TYPE: PRT  
; ORGANISM: *Candida albicans*  
US-11-016-564-3

Query Match 5.6%; Score 100.5; DB 7; Length 331;  
Best local similarity 25.5%; Pred. No. 0.3;

Matches 42; Conservative 23; Mismatches 55; Indels 45; Gaps 8;

QY 9 SDWGVSTFVILETGLDNCNIYANGLMG--VIINITPTDDEGNFVDIDT-----60  
 Db 165 SMFTNIDTKHATVEQ-----TIYINPLRSYAKETTNVISGEGESTI-IDSTIIKRYK 218  
 QY 61 -----LNDNIKIIVDYIDSGIDSGMFTYGNPNEYNT-----IPNS 97  
 Db 219 VGNQNLPSDNRIYDYSEYEDVT-NDIYAQLGNNDVNIINFGINDSPYIIKVISKIDPNK 277  
 QY 98 QSYSLKSENSQITQIKRYVCSNTSRLRTKFSKAV--TTTSGK 140  
 Db 278 DDTYTIQCTVTMQTTINERY-----TGEFTASYDNRTIAFTSSGQ 317

RESULT 5  
 US-10-880-144-6  
 ; Sequence 6, Application US/10880144  
 ; Publication No. US20050287537A1  
 ; GENERAL INFORMATION:

APPLICANT: Svendsen, Allan  
 TITLE OF INVENTION: Pullulanase variants and methods for preparing such variants with  
 TITLE OF INVENTION: predetermined properties  
 FILE REFERENCE: 6072.220-US  
 CURRENT APPLICATION NUMBER: US/10/880.144  
 CURRENT FILING DATE: 2004-06-28  
 NUMBER OF SEQ ID NOS: 40  
 SOFTWARE: PatentIn version 3.3  
 SEQ ID NO 6  
 LENGTH: 829  
 TYPE: PR1  
 ORGANISM: Bacillus acidophilus  
 US-10-880-144-6

Query Match 5.6%; Score 100.5; DB 6; Length 829;  
 Best Local Similarity 21.5%; Pred. No. 0.92;  
 Matches 71; Conservative 44; Mismatches 106; Indels 109; Gaps 15;

QY 11 WTGVSTFVILETGLDNCNIYANGLMG--IGVINTPTDDEGNFVDIDTDLNDNIKI 67  
 Db 551 WTGCT-----SGLSSDQLVTKGQKGLGIV-----FNNIR- 582  
 QY 68 VDIYDSDIDSGDMFTYGNPNEYNTIPN-----SOSYSLKSENSQITQIK 114  
 Db 583 -NGLDGNVFPKSAQGFATGDPNOVNIKRWMSISDFTSAPSETTINYISHDNMTLMDK 641  
 QY 115 RYVSCSN-----TSRLTKSFSKAVTTTS-----GKVISITONSINSSRVINAIDA 161  
 Db 642 --ISASNPNDTQADRIMDELAQAVFTSGVFPMGGBEMLTGKNDS--YNAQDS 696  
 QY 162 TNETDELRTTKRTRENOSYTS-----KSTNSLYVHTWTIPRSLK- 204  
 Db 697 VNQFDSRKQAFENVDPYVSMILHLRDNHPAFMTTADQIKOMLTFDSEFTNTVAPELKV 756  
 QY 205 -LQNMWED-----YNGMTWAGSCYKTAGDGSESTRMLAASIFPPGVDLMDNDI 259  
 Db 757 HANHDMMKAILIYVNTNPKT-----AQTLLTPSGMWTIVGLGNQV 795  
 QY 260 ALSGMAHKSYYNDGTGINQLSFTRIIGKFS 289  
 Db 796 GEKSLGHVNGTVE--VPALS-TILLHGTS 822

RESULT 6  
 US-11-052-554A-92  
 ; Sequence 92, Application US/11052554A  
 ; Publication No. US20050288866A1  
 ; GENERAL INFORMATION:  
 APPLICANT: Sachdeva, et al.  
 TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
 FILE REFERENCE: 30853/40359A  
 CURRENT APPLICATION NUMBER: US/11/052.554A

CURRENT FILING DATE: 2005-02-07  
 PRIOR APPLICATION NUMBER: US 60/589,227  
 PRIOR FILING DATE: 2004-07-20  
 PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
 PRIOR FILING DATE: 2004-02-06  
 NUMBER OF SEQ ID NOS: 763  
 SOFTWARE: PatentIn version 3.3  
 SEQ ID NO 92  
 LENGTH: 2399  
 TYPE: PR1  
 ORGANISM: Helicobacter pylori J99  
 US-11-052-554A-92

Query Match 5.6%; Score 99.5; DB 7; Length 2399;  
 Best Local Similarity 23.0%; Pred. No. 4.1;  
 Matches 75; Conservative 38; Mismatches 128; Indels 85; Gaps 12;

QY 13 GVSTFVILETGLDNCNIYANGLMIGVINTPTDDEGNFVDID-----DVLNDNIKI 66  
 Db 655 GVANNGTLLIGNNTESVNDNGLIWI-----HGFYITGTGFSANITYLTNFK 704  
 QY 67 IVDYIDSDIDSG-----DGWFY-----TG-NPNEYNTIPNSQSYSL 102  
 Db 705 TGEVGNSDGCGANITFKASDNITMDQLNYNNAETVKMIQTGASGHSYTFPATNNISV 764  
 QY 103 LKSENSQITQIKRYVCSNTSRLRTKFSKAVTTTSGKVISITONSINSSRVINAIDA 162  
 Db 765 TDSDFSDMTWGK-----PSFSKNIISFNASPSGFTNGGSSITSTNNSL 811  
 QY 163 NPTDDELK-----TTKETRENOSY-----TSKSTNSLYVHTWTIPRSLKQ 206  
 Db 812 SPTDSLNGALYNLQANSIFNNTOAVRVLYSRGTSFNATLTGLGNTSFLSSQSL 871  
 QY 207 NMRMEDYNNMTWAGSCYKTAGDGSESTRMLAASIFPPGVDLMDNDIALSGMAR 266  
 Db 872 NF-----NGDTLLQ--NANITLGNKQ--AAFNSLTLDNNSMLSLDQSVLANGT 920  
 QY 267 KSYNVDTGIN-----OLSTRIIGK 287  
 Db 921 SAFNQAISLITNGSQAFSSLEFNG 946

RESULT 7  
 US-11-052-554A-91  
 ; Sequence 91, Application US/11052554A  
 ; Publication No. US20050288866A1  
 ; GENERAL INFORMATION:  
 APPLICANT: Sachdeva, et al.  
 TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
 FILE REFERENCE: 30853/40359A  
 CURRENT APPLICATION NUMBER: US/11/052.554A  
 CURRENT FILING DATE: 2005-02-07  
 PRIOR APPLICATION NUMBER: US 60/589,227  
 PRIOR FILING DATE: 2004-07-20  
 PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
 PRIOR FILING DATE: 2004-02-06  
 NUMBER OF SEQ ID NOS: 763  
 SOFTWARE: PatentIn version 3.3  
 SEQ ID NO 91  
 LENGTH: 2902  
 TYPE: PR1  
 ORGANISM: Helicobacter pylori J99  
 US-11-052-554A-91

Query Match 5.6%; Score 99.5; DB 7; Length 2902;  
 Best Local Similarity 17.9%; Pred. No. 5.1;  
 Matches 66; Conservative 63; Mismatches 146; Indels 93; Gaps 14;

QY 15 STFNVILETGLDNCNIYANG--LNNIGVINTPTDDEGNFV-----DI 56  
 Db 145 ATFNLAASSGNSFTSWYPNGHTDVTSPAGTIWNNSEVGNRVGSGAGTGTGATLNLNA 204

```
QY      57 DDVTLNDIKI-----VDYIDGSDIG-----SDGWFT 85
       :|::||:
Db      205 NKVIINSINSAKYTSQVVNVGNANSVTIINSVSLNDGTCSILARVGVCNCTSGSPSYSPK 264
       :|::||:
QY      86 GNPNEYNTIPNSQSYLSLKSNSQTQK-----RYVSCSNTRSLRTKPSFAKTYT 136
       :|::||:
Db      265 GTTNATNTTFSSNSSGFTEENATPFGAKLNGCATTFNKKEPATNTNTAPNSSFTFPKGTS 324
       :|::||:
QY      137 T-SGVISITONSINSTRVINAI---DATNFTELDELRTKETREPNOSYSTSHKASTMSL 192
       :|::||:
Db      325 SPFNAGFBNASSTFFNNQATFQNSSPFNGGTFETFDQNTQSTQHQAQNSSFFSQAATTLKGF 384
       :|::||:
QY      193 VYHTWTIFRS--LKLQWRMEDVYNNGMTWAQSCYYKTG----ADGSESTFRMTLAAGSI 244
       :|::||:
Db      385 ATFEQAFNNSHQLTIQ---ASFNN-----ATFNNTKITIEKXDSFNNT-----SF 429
       :|::||:
QY      245 FPPRGAYDGLMINDNLALSGM-AHKSYNDVTGIGNQSLFRRITCKGFSWYVYNISGLDRGHA 302
       :|::||:
Db      430 NTPVDVNTNMVTISGGVLTSGKNDLKNGCATLDFGSSKITLLQ-----GTFEFLNTLSGESEX 483
       :|::||:
QY      303 VTIIDQYG 310
       :|::||:
Db      484 VTIIINSRG 491
```

```

RESULT 8
US-11-052-554A-94
; Sequence 94, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 94
; LENGTH: 744
; TYPE: PRT
; ORGANISM: Helicobacter pylori J99
US-11-052-554A-94

```

	Query Match	5.4%	Score 96.5;	DB 7;	Length 744;
	Beat Local Smlrlrly	27.2%;	Pred. No. 1.7;		
	Matches	41;	Conservative	22;	Mismatches 63; Indels 25; Gaps 7.
QY	24	GLDNCNIYANGNMGIVILINPTDDEGFVVIDDYLTNDINKIYDIYIDGSDIDG-SDGW	82		
	:: :	:	:	:	:
Dd	213	GNMNCSSQVTGVNN-----QKDGTKTKITOTIDGKSVTITTISSKYVD-----SPADENTTGV	263		
QY	83	PYTGNPNENYNTIPNSQSYSLSKENSQITQIKR---YVCSNTSLRTKFSFSAKYTTTSG	139		
	:: :	:	:	:	:
Dd	264	SYTEITNKLEGPVDS-AQLALAQASTLINTINNACPYFHASNSEANAPKFS-----TTTG	318		
QY	140	KV-----ISTONSINSRRVYNIAIDATN	163		
	:	:	:	:	:
Dd	319	KICGAFFSEISAIQKMITDADELVNQTSVIN	349		

RESULT 9  
US-11-052-554A-55  
; Sequence 55, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

```

? FILE REFERENCE: 30853/40359A
?
? CURRENT APPLICATION NUMBER: US/11/052,554A
?
? CURRENT FILING DATE: 2005-02-07
?
? PRIOR APPLICATION NUMBER: US 60/589,227
?
? PRIOR FILING DATE: 2004-07-20
?
? PRIOR APPLICATION NUMBER: IN 173/DEL/2004
?
? PRIOR FILING DATE: 2004-02-06
?
? NUMBER OF SEQ ID NOS: 763
?
? SOFTWARE: PatentIn version 3.3
?
? SEQ ID NO 55
?
? LENGTH: 807
?
? TYPE: PR1
?
? ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-55

```

[illegible]

```

RESULT 10-657-1088
US-10-467-657-1088
/ Sequence 1088, Application US/10467657
/ Publication No. US20050260581AI
/
/ GENERAL INFORMATION:
/ APPLICANT: CHIRON SPA
/ APPLICANT: FONTANA Maria Rita
/ APPLICANT: PIZZA Mariagrazia
/ APPLICANT: MASIIGNANI Vega
/ APPLICANT: MONICI Elisabetta
/ TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
/
/ FILE REFERENCE:
/ CURRENT APPLICATION NUMBER: US/10/467,657
/ CURRENT FILING DATE: 2003-08-11
/ PRIOR APPLICATION NUMBER: GB-0103424.8
/ PRIOR FILING DATE: 2001-02-12
/ NUMBER OF SEQ ID NOS: 9218
/ SOFTWARE: SeqwIn99, version 1.04
/ SEQ ID NO 1088
/
/ LENGTH: 1468
/
/ TYPE: PRT
/
/ ORGANISM: Neisseria gonorrhoeae
US-10-467-657-1088

```

```

Query Match          5.4%; Score 95.5; DB 6; Length 1468;
Best Local Similarity 22.6%; Pred. No. 4.7;
Matches 63; Conservative 36; Mismatches 109; Indels 71; Gaps 15;
OY 9 SWMTGVSFTPNVILLEGDNCNIRYANGLMIGVITINPTDEGNAFVDD-----V 59

```

```

Db 729 SDWTGLSCRE--KTIYDD-----KVIAISKTDIRGN-VSLADHAHLNLTGLA 774
Qy 60 TLADNRIKYDIYDSDIDSDGDMFYTNPNREYNTIP-SQSYSLKENSQITQIKRYVS 118
Db 775 TLNGLNIS-----ACGDHYTYTRATQNGNLSLGNAAQT-FNQATL 815
Qy 119 CSNTRSRKTSFSAKVTTSKGVISITONS-INSSRVINAIIDNTFTDELRTKTRF 177
Db 816 NGNTSASDNASFNLSNNAVONGSLTLDNAANVSHALNG-----NVSADKAVFHF 866
Qy 178 ENOSYTSKHSSTNSLYVH-----TWTPRSLKIQMRMEDYNNGWTAQSCYY--KTGADG 231
Db 869 ENSRFTKISGKOTATLHLDSEWTLPSGTELGSL---NIDNATITLNSAYRHDAQAQT 925
Qy 232 GS-----ESTRWLAAGSIFPPGNYDGLMDNDIALSG 263
Db 926 GSADAPRRRSRRSL--SVTPPTSASERF--NLTAVNG 960

```

## RESULT 11

```

US-11-052-554A-17
; Sequence 17, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 17
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-17

```

Query Match 5.3%; Score 95; DB 7; Length 980;  
 Best Local Similarity 19.8%; Pred. No. 3.2;  
 Matches 68; Conservative 55; Mismatches 108; Indels 112; Gaps 18;

```

Qy 5 EDMSDWTGVSTFNVILETGLDNCNY-----ANGLMIGVITINPTD 48
Db 147 DDVYEHYTYNGITNNADTHV--VDVYDMGTAITLDEVDLSTINSHVAGITLT----- 199
Qy 49 DEG-NFVDIDVTL-----NDNIKIVD--YIDGSDID-GSDGMF-YTGNPNRYNT 93
Db 200 -QGWEYEDIDNTVSTGVSSEVENNTITVDSFTVTSGMTDEGTGTFMFGHTGASNY-- 256
Qy 94 IIPNSQSILKSENSQITQIKRYVSCNTRSRKTSFSAKVTTSKGVISITONSINSR 153
Db 257 -----SNTLTADVAIAAANPYADNMQTTY-TLDNST 289
Qy 154 VVINAIIDATNFTDELRTKTRPENOSYTSKHSSTNSLYVHTWTPRSLKIQMRMEDY 213
Db 290 LMGDVVPSSNFDEN-----FFPGANSYRPAADDVUTNGMDGDRNDV-----TL 334
Qy 214 NNG--WTMAQSCYYKTGADG-----GSESTRWL--AAGSIFPPG-----NY 250
Db 335 NNGSKWGAAMSYMVEDDEGSDYGYAVGTGATATLIDIAANSLWPSSTVGVNDINTQY 354
Qy 251 D--GLMDNDIALSGMAHKSYN-----VDTGINQLSFRIITGK 287
Db 395 DENGHIVNEVYQGLFNVTILNGSEWDYTKSSLDITLINSNG 437

```

## RESULT 12

```

US-11-013-759-11
; Sequence 11, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; FILE REFERENCE: 1038-921MIS:jb
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-11

```

Query Match 5.3%; Score 94; DB 7; Length 2314;  
 Best Local Similarity 22.0%; Pred. No. 11;  
 Matches 68; Conservative 44; Mismatches 103; Indels 94; Gaps 17;

```

Qy 44 ITPTDEGNFVDIDV-----TLNDRIKYDIYDSDIDSDGDMFYTNPNREYNTIP--- 95
Db 748 ITSDDEKSKAASIGDILNTGFNLKNNNSNGVF-----STNTGVFID 790
Qy 96 -NQSYSLKSENSQITQIKRYVSC-----SNTSRKTSFSAKVTTSKGVIS 143
Db 791 GNAITAKVYDEFTNQSKYVDVAVDEKTELTGDNKTKIKGVKTTTLTITNANGKA-- 848
Qy 144 ITONSINSSRVINAIIDAT--NFTDELRTKTR-----RFENOSYTSKH 186
Db 849 -TNFSTTDNALVNAADIAENANTLAKETHTTGATADLTQTRKVKDQATDEITIVGK 907
Qy 187 SST-NSLYVHTWTPRSLKIQMRMEDYNNGWTAQSCYYKTGADG--GSESTRWLA 241
Db 908 DGTQNGKTVNT-----LKLKG-----ENGLTVA-----TKDGTVTFGINTQSLKA 949
Qy 242 GSIFPPGNYDGLMDNDIALSGMAHKSYNVDTGINQLSFRIITGKPSWYINISGLDRGH 301
Db 950 GD-STLNDGSLSKIPAS-----NEQIVGADGVKFAK-VDKNS--STGID-GT 995
Qy 302 AVTIIDQYV 310
Db 996 SRITKDIQ 1004

```

```

RESULT 13
US-10-416-047-14
; Sequence 14, Application US/10416047
; Publication No. US20050266512A1
; GENERAL INFORMATION:
; APPLICANT: Buckley
; TITLE OF INVENTION: Detection of Proteases and Screening for Protease Inhibitors
; FILE REFERENCE: 63198-133
; CURRENT APPLICATION NUMBER: US/10/416,047
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/247,160
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: PCT/CA01/01561
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Aeromonas hydrophila
US-10-416-047-14

```

Query Match	5.2%	Score 93.5;	DB 6;	Length 486;
Best Local Similarity	19.1%;	Pred. No. 1.8;		
Matches 66;	Conservative 41;	Mismatches 108;	Indels 131;	Gaps 15;

```

Qy      8 MSWM---TGSTNNVLLEGLD-----CNIYANGNMI---GVIIINTFPD---DEGNVDI 56
Db      65 MGQWQISGLANGVIMPGVNGEIKRGTASNWCYPTPVGRIPTSLALDIPDGEDVD 124
Qy      57 DDVTLNDN---IKIYDID-----GSDID---GSDMPFTGNPN----- 89
Db      125 QMRLVHDSANFIKPTSYLAHYGVYMWGNSQYVGEIMDVTBDGDCWVIRGANDGCCDG 184
Qy      90 -----EXNTIPNSQSYLSKSENSQITQ-----IKRY 116
Db      185 YRCGDKTAIKVNSFANLDPDPSFKGADVTSQDRQLVKTVMGAVANDSDTPOSGYDVLTRY 244
Qy      117 VGSQNSRLRTSFSFAKYTTTS---GKIYSTQNSI---NSRNVINAI DATNFTDDEL 169
Db      245 DTAIINSKNTNYTGLSEKVTYTKNKKFMPVLEGTEQLSIEIAHQSMASONGSSTTSSLSQV 304
Qy      170 RTTKER-----FENOSYTSH-----KSGTNSLYH-----TWT- 198
Db      305 RPTVPARSKPIVKIELYKADISYRPFEPKADVSDYDTLSEGLRMCAGNAVYTHPDNRPMWNH 364
Qy      199 -----IPPSLKQNRMEEDYNNGMTWQS 222
Db      365 TPVIGPYDKASIRYQMDKRIITPGEVTKMDNMWITIQONGSLSTQN 410

```

```

RESULT 14
US-11-089-551A-33
: Sequence 33, Application US/11089551A
: Publication NO. US20050266242A1
: GENERAL INFORMATION:
: APPLICANT: Lindquist et al.
: TITLE OF INVENTION: ELECTRICAL CONDUCTORS AND DEVICES FROM PRION-LIKE PROTEINS
: FILE REFERENCE: 30554/40025A
: CURRENT APPLICATION NUMBER: US/11/089,551A
: CURRENT FILING DATE: 2005-03-24
: PRIORITY APPLICATION NUMBER: US 60/559,286
: PRIORITY FILING DATE: 2004-03-31
: NUMBER OF SEQ ID NOS: 65
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 33
: LENGTH: 710
: TYPE: PR1T
: ORGANISM: Saccharomyces cerevisiae
US-11-089-551A-33

```

Query Match	5.2%	Score 33.5	DB 7	Length 710
Best Local Similarity	22.8%	Pred. No. 2.8		
Matches	41	Conservative	31	Mismatches 85; Indels 23; Gaps 5
QY	70	YIDSGDIDGSDGWFYTGPNENYNT-----IPNSGSYSLSLNSNSQITQIKRYVCSNTS	123	
DB	6	YNGNSNVDTG----YDRNDTRMTNWARSVRVDKGRSSSTSKGYRRTRAGSDPTTNSA	61	
QY	124	RLRTYSFSAKTTTSGKVIISTQNSINSRVYNAIDATNPFDELRTTKERFENQSYT	183	
DB	62	KHSHKSKTIVVVTST-----STDNSITTYARVSDSTVATIS-----STTTRTRTNNTVVS	110	
QY	184	SHKSTNSLVIYHTTITPRSLKIQNRMWEDYNNGMWMAQSCYYKTGADGSESTRMLAAGS	243	
DB	111	STASSSTTDVGNATSAANNSANASNTSSSDYATSYRKSTNDNTTIA--NSKGNMWSAKON	168	

RESULT 15  
US-11-052-554A-9  
; Sequence 9, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.

```

1  TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
2  ;
3  ; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
4  ;
5  ; FILE REFERENCE: 30853/40359A
6  ;
7  ; CURRENT APPLICATION NUMBER: US/11/052.554A
8  ;
9  ; CURRENT FILING DATE: 2005-02-07
10 ;
11 ; PRIOR APPLICATION NUMBER: US 60/589,227
12 ;
13 ; PRIOR FILING DATE: 2004-07-20
14 ;
15 ; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
16 ;
17 ; PRIOR FILING DATE: 2004-02-06
18 ;
19 ; NUMBER OF SEQ ID NOS: 763
20 ;
21 ; SOFTWARE: PatentIn version 3.3
22 ;
23 ; SEQ ID NO 9
24 ;
25 ; LENGTH: 1579
26 ;
27 ; TYPE: PRT
28 ;
29 ; ORGANISM: Escherichia coli 0157:H7
30 ;
31 ; US-11-052-554A-9

```

	Query Match	5.2%; Score 92.5; DB 7; Length 1579;	
	Best Local Similarity	18.9%; Pred. No. 9;	
	Matches	68; Conservative	67; Mismatches 121; Indels 103; Gaps 18;
Qy	3 LREDSMSDWT--GVSTFNVILETGLDNCNI---YANGLNMI-GVLIINIRP--TDDEGN	52	
Db	415 VQDITLSDMTDNBGSYTQILTAGTSOSVTLTPQINSASAKESIVNIPVWSSRHS	474	
Qy	53 FVDIDDT--LNDNIKI-VDYIISGDDIDSGMFTYGNPNRYN-----TIPISSO--	99	
Db	475 SITIDNVSYYAGDIKRAVELKDSON-----QPVAYQKEELVKAATVENSKGA	523	
Qy	100 -----YSLLKSENSQITQIKRVYSCSN-TSRLRTKSFSKAVTTTSGKV--ISIT	145	
Db	524 TIYWHEEOPGYAANYPAYKOGTALRAQLSLHNNAPAQSHIYNLEANOQRARVATTBAT	583	
Qy	146 QNSINSKRVIN-AIDATNFETDELRTKTETREMOSGTSH-----KSTNSLYVHTWT	198	
Db	584 NNDVYADKKTEFNTLITLVTDSDNPL-TNHQVFTEFKGSAEFVEPPPOONTDAYGVATIN	642	
Qy	139 IPRSLIKONMWEDYNNGTAAQSCTYYTGAGDGSESRMLAAG-----	242	
Db	643 MVSQVAENITISATLPNG-PSORIIAFVDBSSTPKRKQVLVADPTTIAGNSQGSTLTLA	700	
Qy	243 -----SIIPPGNYDGLMYLDNIIA-----LSGMHKHSYNDPTGI	275	
Db	701 IITDFHNNPLODKMKNEVPAC--GSQDLNTATYITDQSIGYARVHLTSSKAGSYSTDALS	756	

Search completed: January 30, 2006, 09:53:18  
Job time : 3.66268 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: January 30, 2006, 09:41:43 / Search time 5.82461 Seconds

(without alignments)  
5517.344 Million cell updates/sec

Title: US-09-889-874a-22

Perfect score: 1784

Sequence: 1 FTLRSDMSDWTGVSTFNV.....ILFHAGYNSDPYLSSTIVY 334

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Listing first 45 summaries

## Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	7.0	1092	2 T30214	fibrinogen-binding
2	122.5	6.9	990	2 E90433	conserved hypothe
3	117.5	6.6	640	2 B32935	hypothetical prote
4	115.5	6.5	1671	2 S71628	sensory transducti
5	111	6.2	369	2 F95107	glutamate 5-kinase
6	111	6.2	376	2 H97975	glutamate 5-kinase
7	109	6.1	807	2 B71605	hypothetical prote
8	109	6.1	1449	2 T30857	glucosyltransferas
9	109	6.1	1449	2 T30552	glucosyltransferas
10	108	6.1	626	2 T01485	probable polygalac
11	107.5	6.0	343	2 E70149	hypothetical prote
12	107.5	6.0	1038	2 H90053	hypothetical prote
13	107	6.0	1072	2 A86827	hypothetical prote
14	107	6.0	1114	2 UH0284	125k surface anti
15	107	6.0	1449	2 B81963	19A-specific serin
16	106	5.9	639	2 A32935	protein p1 - Entam
17	105.5	5.9	439	2 G97336	probable S-layer p
18	105	5.9	507	2 A71622	hypothetical prote
19	105	5.9	635	2 S57714	capB protein - Clo
20	105	5.9	667	2 E69733	PSX prophage ORF
21	105	5.9	1431	2 A45866	extracellular neur
22	104	5.8	836	2 D97182	extracellular neur
23	104	5.8	1306	2 S25370	MSB2 protein - Yea
24	103.5	5.8	321	2 A69088	conserved hypothe
25	103.5	5.8	934	2 T39941	hypothetical prote
26	103.5	5.8	1031	2 A29839	RAM2 protein - Yea
27	103.5	5.8	1881	2 H95076	zinc metalloprotei
28	103	5.8	497	2 T47715	hypothetical prote
29	102.5	5.7	531	2 B96966	beta-mannanase Man

30	102.5	5.7	1148	2 S72635	exo-poly-alpha-gal
31	102	5.7	934	2 B29838	paraportal crystal
32	102	5.7	1224	2 E71611	hypothetical prote
33	102	5.7	2894	2 C64474	hypothetical prote
34	101.5	5.7	557	2 JC5487	cellulase (EC 3.2.
35	101	5.7	381	2 T28378	ORF MSV217 SCG gen
36	101	5.7	870	2 S27514	mosquitocidal toxi
37	101	5.7	903	2 T20804	hypothetical prote
38	101	5.7	1136	1 USB881	paraportal crystal
39	101	5.7	1457	2 D81019	adhesion and penet
40	101	5.7	3194	2 D71917	toxin-like outer m
41	101	5.7	3972	2 S75251	hypothetical prote
42	100.5	5.6	1649	2 C86822	hypothetical prote
43	100	5.6	1577	2 A35140	hemolysin A precu
44	100	5.6	2329	2 T28125	hypothetical prote
45	99.5	5.6	962	2 AG2444	hypothetical prote

## ALIGNMENTS

RESULT 1	
T30214	fibrinogen-binding protein - Staphylococcus epidermidis
C:Species: Staphylococcus epidermidis	
C>Date: 22-Oct-1999 #sequence: revision 22-Oct-1999 #text: change 09-Jul-2004	
C/Accession: T30214	
R:Nilsson, M.; Frykberg, L.; Flock, J.I.; Pel, L.; Lindberg, M.; Guss, B.	
Infect. Immun. 66, 2666-2673, 1998	
A>Title: A fibrinogen-binding protein of Staphylococcus epidermidis.	
A:Reference number: Z20781; PMID:98261511; PMID:9596732	
A:Accession: T30214	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-1092 <N1>	
A:Cross-references: UNIPROT:O70022; UNIPARC:UPI00000AFCA6; EMBL:Y17116; NID:e1296734; PIR	
Query Match	7.0%; Score 124; DB 2; Length 1092;
Best Local Similarity	21.0%; Pred. No. 0.58;
Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;	
Qy	9 SDWTGVSTFNVILETGLDNCNIYANGIMIG--VIINPTDDEGNFVDVDT----- 60
Db	439 SMTNIDTKHYEQ-----TIINPLRYSAKTNVINSNGDEGSI-IDSTIIIVYK 492
Qy	61 -----LNDIKIVDYIDSGIDSGWFYTGPNEXNT-----IPNS 97
Db	493 VGDNQMLPDSNRIVDYSEYEDVT-NDVYAQLGNNDVNIINFGNIDSPYIIKVISKDPNK 551
Qy	98 QSYSLKSENSQITQIKRYVSCNTSRIRKFSYSAKY--TTTSG-----KVISI-- 144
Db	552 DDYTTIQQVTVMTTINEXY-----TGEFRASYDNTIAFTSSGQGGDLPPERTYKIGD 606
Qy	145 -----TQINSRVRVINAIDATNPTDELRTTKTRFENSGSYSHKSTNSL-Y 193
Db	607 YWEDVDKQIQNTNNEKPLSNVLTLYTPDG---TSKSVRTDEBGKYPDGLKNGLT 663
Qy	194 VHTWTIPRLIKLQNMWMBDY-----NNGMTWAQSCYYTGADGSEST 236
Db	664 KTFETP-----EGYPTLKSGTNPALDSRGNVWVINGDDMTIDSGFYQTP 713
Qy	237 RMLAASIFPPGNYDGLMDNDIALSGMAHKSIVNVTGINSFTNIIIGKFSWVNIISG 296
Db	714 KYSLGNYWYDNTKKDIIQGDDEKIGISV--KVTLKDENGNIISTTTDENG---KYQPDN 768
Qy	297 LDRGHAVIITDQ 308
Db	769 LNSGNYIVHFDK 780
RESULT 2	
E90433	conserved hypothetical protein [imported] - Sulfolobus solfataricus

C:\Species: Sulfolobus solfataricus  
C:\Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
A:\Accession: B30433  
C:\Species: Entamoeba histolytica  
C:\Date: 20-Dec-1989 #sequence\_revision 20-Dec-1989 #text\_change 09-Jul-2004  
A:\Accession: B32935  
R:\Tannich, E.; Horezmann, R.D.; Knobloch, J.; Arnold, H.H.  
Proc. Natl. Acad. Sci. U.S.A. 86, 5118-5122, 1989  
A:\Title: Genomic DNA differences between pathogenic and nonpathogenic Entamoeba histolytica  
A:\Reference number: A32935; MUID:8929655; PMID:2544890  
A:\Accession: B32935  
A:\Status: preliminary; not compared with conceptual translation  
A:\Molecule type: mRNA  
A:\Residues: 1-640 <TAN>  
A:\Cross-references: UNIPROT:P20301; UNIPARC:UPI0000125CB8

Query Match      6.9%; Score 122.5; DB 2; Length 990;  
Best Local Similarity 20.8%; Pred. No. 0.66;  
Matches 64; Conservative 45; Mismatches 99; Indels 99; Gaps 13;

Dy      36 NMIGVINITPTDEGNFVIDD-----YIANDNIKIIVYIDGSDIDSGDW 82  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Db      728 NLVINISLKGKTSN--NVIIYNQPVSLKAHYLOLYINLSINIKALINSTLTSPGW 765  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Oy      83 FYTGNP-----NEVYTPINSQSYSLL-----KESNSQITQIKKYSCSNTSRILRTK 128  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Db      786 YNAVTLPISFIWMTYYISNNTRYIILSLPFPFTVNRSLTVKVTLLKEYL----- 834  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Oy      129 SPSAKVTTTSGKVSIITONSINSRRVYINAIDANFDDELRTTKERFRFNQSYTHSKSS 168  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Db      835 -----VYNEBILTLTIANRLTSEIWPA-----GQTLLPKRYVNISNNRIFYNTS 882  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Oy      189 TNSLVHTWTIPRSKLONMMRWEDYNNNGTWAGSCYYTGTGADGGSSESTRWLAAAS----- 243  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Db      883 S---YLNIITOPTSLNVK-----PIIEYIVTDGNSE---WLPGRSVTLTT 922  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Oy      244 ----TFPPGNYDGLMDNDIALSGMAHKSYNVDGTGINQLSFTRIIGKFSWVYNISGLDR 299  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Db      923 QSVPIEYGGKWEK-----SYAVNSGV-AITVNOPIETEF--VKNIINGSFV 964  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Oy      300 GHAVIII 306  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Db      965 GSVALILI 971

RESULT 3  
B32935  
hypothetical protein NP1 - Entamoeba histolytica (fragment)  
C:\Species: Entamoeba histolytica  
C:\Date: 20-Dec-1989 #sequence\_revision 20-Dec-1989 #text\_change 09-Jul-2004  
A:\Accession: B32935  
R:\Tannich, E.; Horezmann, R.D.; Knobloch, J.; Arnold, H.H.  
Proc. Natl. Acad. Sci. U.S.A. 86, 5118-5122, 1989  
A:\Title: Genomic DNA differences between pathogenic and nonpathogenic Entamoeba histolytica  
A:\Reference number: A32935; MUID:8929655; PMID:2544890  
A:\Accession: B32935  
A:\Status: preliminary; not compared with conceptual translation  
A:\Molecule type: mRNA  
A:\Residues: 1-640 <TAN>  
A:\Cross-references: UNIPROT:P20301; UNIPARC:UPI0000125CB8

Query Match      6.6%; Score 117.5; DB 2; Length 640;  
Best Local Similarity 21.9%; Pred. No. 0.83;  
Matches 73; Conservative 45; Mismatches 121; Indels 95; Gaps 18;

Oy      25 LDNCNIYANG---LNMIGVINITPTDEGNFVIDDITLDNDIKIIVYIDGS-----DI 76  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Db      17 LDNIRMSNDMRPIDSIGDLGLQT-----QPYSIDTFKISSPGIMVIYLRDT 67  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Oy      77 DSGDGWFYT---GNPN--EYNTIPNSQSYSLLKSENSQITQIK---RVYSCNTSRILR 126  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
Db      68 TFTNSFYVTFYVNVGASAIINYNTTNMBENWSVLKNAFGVAEIRTPGNRLVFTSR--HIR 125  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

```

Oy 127 TKSSBAKTTTSGKIVTSITONSINSRVYINAIIDANFTDDELRTKERTFENOSYSHK 186
Db 126 NLEDAQYISDWLWKISISNAVLENTPI---TLNF-DQRYDAGANAFAVDFKFTQHP 180
Oy 187 SSTNSLYVHTWTIPRSLKLQNRWEDYNNGMTWAQSCYYKTGADGSESTRW--LAAGSI 244
Db 181 SD-----WASCSYAK---EGLINSQNWGPHLHMNH 207
Oy 245 PPPGNY--DGLMLNDIALSGMAHKSYNVDTGINQLSFTRIIQ---KGFs-WYVYNISGLD 298
Db 208 HMQGYLRGKWN---GIKEPGEERTNNWTSINYLITVNIAGHRQGLSGMNY----- 256
Oy 299 RGHAVIITIDYGNKRIILFHAGYENSDDPYLSSSI 332
Db 257 -----VSDGIISTYKIL--NGENDQPHLRSTV 281

RESULT 4
S71628
sensory transduction histidine kinase dokA - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #ext_change 09-Jul-2004
C:Accession: S71628; S78068
C:Restructer, S.C.; Noegel, A.A.; Oehme, F.; Gerisch, G.; Simon, M.I.
EMBO J. 15, 3880-3889, 1996
A>Title: The hybrid histidine kinase dokA is part of the osmotic response system of Dicty
A:Reference number: S71628; MUID:96324396; PMID:8670893
A:Accession: S71628
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1670 <SCW>
A:Cross-references: UNIPROT:Q23901; UNIPARC:UPI000017CBFD; EMBL:X96869
A:Experimental source: strain AX2, substrain 214
R:Restructer, S.C.; Noegel, A.A.; Oehme, F.; Gerisch, G.; Simon, M.I.
submitted to the EMBL Data Library, March 1996
A:Description: The hybrid histidine kinase dokA is part of the osmotic response system of
A:Reference number: S78068
A:Accession: S78068
A:Molecule type: DNA
A:Residues: 1-149,'E',151-219,'TRVLKLIQSTNNMITYV',238-1671 <SCW>
A:Cross-references: UNIPARC:UPI000080237; EMBL:X96869; NID:91237201; PIDN:CAA65612.1; P
C:Genetics:
A:Gene: dokA
C:Function:
A:Description: modulates cell response to changes in osmolarity; involved in spore forma
C:Keywords: phosphoprotein; signal transduction
F:1520-1629/Domain: response regulator homology #RRH2>
F:1568/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 6.5%; Score 115.5; DB 2; Length 1671;
Beet Local Similarity 22.3%; Pred. No.4.3; Mismatches 91; Indels 99; Gaps 11;
Matches 64; Conservative 33;

23 TGLDNCNIYANGIMIGVINITPTDDEGNFVDIDVTIANDIKIVYIDSGDSDGW 82
1347 TMLDLESSEKSTENQDIIIV---GSDKSTFVQ-----QILDVYTEKV-----I 1387
Oy 83 FTTGAPN-EYNTIPNSQSYSL-----LKSNSQITQIKRYVSCSNT 122
Db 1388 LYGVDPNKSN-EENSKVSYLVTPITPSKLISSILSKNKLKSNISFLTTNNNTNNT 1445
Oy 123 SKLRFSFAKAYTTTSGKIVTSITONSINSRVYINAIIDATNFTDDELRTKERTFENOSY 182
Db 1446 NNIEKNS-----SIDSPLSITTSISSIITPILISNNLIDAN----- 1481
Oy 183 TSHKSTNSLYVHTWTIPRSLKLQNRWEDYNNGMTWAQSCYYKTGADGSESTRWLAAG 242
Db 1482 -NNNNNNNSILVS-----NNG-----GVDGNNVPELTITTI 1511
Oy 243 STPPGNYGLMLNDIALSGMAH---SYNVDTGINQLSFTRIIIG 285
Db 1512 QQSOPKRYLVAEDNDINIKVVRQLEKIGYTAIVGINGKALBIIIG 1558

```

## RESULT 5

glutamate 5-kinase [imported] - Streptococcus pneumoniae (strain TIGR4)

C/Species: Streptococcus pneumoniae  
C/Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C/Accession: F95107

R/Retelin, H.; Nelson, K.R.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidom, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple, neon, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001

A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A/Reference number: A95000; PMID:21357209; PMID:11463916

A/Accession: F95107  
A/Status: preliminary  
A/Molecule type: DNA

A/Residues: 1-369 <KUR>

A/Cross-references: UNIPROT:Q97R95; UNIPARC:UPI0000051644; GB:AE005672; PIDN:AAK75055.1;  
A/Experimental source: strain TIGR4

C/Genetics:

A/Gene: SP0931  
C/Superfamily: glutamate 5-kinase

Query Match 6.2%; Score 111; DB 2; Length 369;

Best Local Similarity 25.0%; Pred. No. 1.2;  
Matches 45; Conservative 34; Mismatches 73; Indels 28; Gaps 8;

QY 31 YANGLMIGVITN--ITPTDEGNFVDIDVTLANIKIVYIDG-----SDIDG 78

DB 113 YKHAQALSTLNRGAIPINENDSVYIDELKVGDNLTLSAQVAAVQADLVFLDVG 172

QY 79 SDGWFYTGPNENYNTIPNSQSYSLKSENSQITQIKRYVSCSN--TSRLRTKSFSAKYTTT 137

DB 173 ----LYTGNP---NSPRAKRLRIETINREIIDMAGAGSSNGTGMLTKIRAAATATE 225

QY 138 SGKVISITONSINSRVINAIDATNFTDDE---LRTKETREPNOSYTSKSTSLVY 194

DB 226 SGVPVVIC-SLKSND---SMIEAETEDGSYFVAQEKGLRQKQWLAFYASQGSIMW 280

## RESULT 6

glutamate 5-kinase (EC 2.7.2.11) [imported] - Streptococcus pneumoniae (strain R6)

C/Species: Streptococcus pneumoniae  
C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C/Accession: H97975

R/Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M, Y. P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001

A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A/Reference number: A97872; PMID:11429245; PMID:11544234

A/Accession: H97975

A/Status: preliminary  
A/Molecule type: DNA

A/Residues: 1-376 <KUR>

A/Cross-references: UNIPROT:Q8DQ61; UNIPARC:UPI0000083527; GB:AE007317; PIDN:AAK99636.1;  
C/Genetics:

A/Gene: proB

C/Superfamily: glutamate 5-kinase  
C/Keywords: phosphotransferase

Query Match 6.2%; Score 111; DB 2; Length 376;

Best Local Similarity 25.0%; Pred. No. 1.2;  
Matches 45; Conservative 34; Mismatches 73; Indels 28; Gaps 8;

QY 31 YANGLMIGVITN--ITPTDEGNFVDIDVTLANIKIVYIDG-----SDIDG 78

DB 120 YKHAQALSTLNRGAIPINENDSVYIDELKVGDNLTLSAQVAAVQADLVFLDVG 179

QY 79 SDGWFYTGPNENYNTIPNSQSYSLKSENSQITQIKRYVSCSN--TSRLRTKSFSAKYTTT 137

DB 180 ----LYTGNP---NSPRAKRLRIETINREIIDMAGAGSSNGTGMLTKIRAAATATE 232

QY 138 SGKVISITONSINSRVINAIDATNFTDDE---LRTKETREPNOSYTSKSTSLVY 194

DB 223 SGVPVVIC-SLKSND---SMIEAETEDGSYFVAQEKGLRQKQWLAFYASQGSIMW 287

## RESULT 7

hypothetical protein PFB0850c - malaria parasite (Plasmodium falciparum)

C/Species: Plasmodium falciparum  
C/Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C/Accession: B71605

R/Gardner, M.J.; Telletlin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; .; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.; Science 282, 1126-1132, 1998

A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A/Reference number: A71600; PMID:99021743; PMID:9804551

A/Accession: B71605

A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA

A/Residues: 1-807 <GAR>

A/Cross-references: UNIPARC:UPI000017B60F; GB:AE001420; GB:AE001362; NID:G3845287; PIDN:;  
A/Experimental source: clone 3D7

C/Genetics:

A/Gene: PFB0850c

Query Match 6.1%; Score 109; DB 2; Length 807;

Best Local Similarity 19.9%; Pred. No. 4.7;  
Matches 68; Conservative 59; Mismatches 122; Indels 92; Gaps 13;

QY 20 ILETG-LDNCNIYAN-----GLNMIGVITNPTDEGNFVDID--DVTL 61

DB 44 INSGHMSNDINIKNQEKKKKKKKKKTKVAINNTHINHTNDKNNQODINKPEVIE 103

QY 62 NNDIKIVYIDGSDIGSDGWFYTGPNENYNTIPNSQSYSLKSENSQITQIKRYVSCSN 121

DB 104 RDNINIKNDTNNIILDS--YNEEGENNRNDNNNNNNNNNNNNNNNNNN--SCSN 156

QY 122 TSRLRTKSFSAK-----VTTTSGK-----VISITONSINSRVINAIDA 161

DB 157 NYGLKKKITLKRNDIKDEGYNNENITTLNNKNLKNNNNNYNNRRNNNNNNNNNN 216

QY 162 TNFTDELRTYK-----TRFENOSYTSKSTSLVYHTWTLPRLSKLQWRW 210

DB 217 NNCSEKTELEQREKEYNKIRARIFSNFNKKQKVQKTEQNNLN-HTY-----LNNNII 268

QY 211 EDVNNG---WTMAQSCYKTGADGSESTRMLAAGSIFPGAVDGLMDNDLISGAHK 267

DB 269 NNINNDNOYAYINNYH-----YHNNSNTHIYRNNIPICNINNH 310

QY 268 SYAVDT-----GINQSFTRIIGKGFVWYISGLDRGH 301

DB 311 APNIEKLNPPYTHDHIAVTNMYSTQKNNMKTKQIGH 351

## RESULT 8

glucosyltransferase - Streptococcus salivarius

C/Species: Streptococcus salivarius  
C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T30857

R/Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995

A/Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for prn  
A/Reference number: Z20909; PMID:95122197; PMID:7822030

A/Accession: T30857

A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA

A/Residues: 1-1449 <SIM>  
A/Cross-references: UNIPROT:Q55264; UNIPARC:UPI000008166F; EMBL:L35495; NID:G662378; PID

C;Genetics:  
A;Gene: gtf1

Query Match	Score	DB 2	Length
6.13	109	1449	

```

OY      17  FNVLIELETDJCN--IYANGIMMIVIIIN--ITPTDEGN--VIDIDVTLNDIKI
Db      459  FELLANDIDNSNPAVQAEQJLNMWLMHYIMNIGITIGGSEDENDGVAVDAVDVNDLLQI
OY      68  V-DYID--GSD-----IDSGDCFFYTGNPMEN-----IIPNSQ-----YSL 102
Db      519  ASDVFYKARYGADQSOQDAIKHLSTILEASH--NDAYNEDTYGQAQJLPMDDPMLHALVSL 576
OY      103  LK--SENSQITQI-----KRYVCSNSTRSLTKSF--SAKVTTTSGKVISITONSIN 150
Db      577  LRPIGNRGVEPLISNLSNDRESEGSNKRMANVAFVRAHDEVSQIIQOI--KNEIN 633
OY      151  SSRVIVNADATNFPDDLEKT-----TKETRENSQSYSHK--SSNLSLYHTWTPR-- 201
Db      634  POST-----GNTFTLDEWKKAFEIYNKDMRSANQYQYUNIPSAALMLTHKDVPRVY 687
OY      202  -----SIKLONMRWEDYNNGMTVA---Q 221
Db      688  YGDWVTTDDQGYWAQKSPYYDAIETILLKGRIRYAAGQDQKV--NYIGYNTNGMDAGVLT 746
OY      222  SCYITKTA---DGSSESTRMLAAGSIFFPGNYVDGJMLNDLALS--GMHKS 268
Db      747  SVRYGTGANSASDGTGAETR--NQGNAVIVSNOPARLTSLSNLTINNGAHRN 796

```

## RESULT 9

glucosyltransferase N - Streptococcus salivarius (fragment)  
C|Species: Streptococcus salivarius  
C|Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C|Accession: J30552  
R|Jaffe, R.I.  
submitted to the EMBL Data Library, February 1998  
A|Description: Streptococcus salivarius V1477 gtfN.  
A|Reference number: Z20854  
A|Accession: J30552  
A|Status: preliminary; translated from GB/EMBL/DBJ  
A|Molecule type: DNA  
A|Residues: 1-1449 <JAF>  
A|Cross-references: UNIPROT:O66542; UNIPARC:UPI000000B10FD; EMBL:AF049609; NID:g2935545  
C|Genetics:  
A|Gene: gtfN

Query Match	6.1%	Score 109	DB 2	Length 1449
Best Local Similarity	22.7%	Score No. 10		
Matches	80	Conservative	45	Mismatches 113; Indels 114; Gaps 22
Qy	17	FNVLILETGLDNON--ITYANGLMIGIYIN--ITPDDENF----	VIDDYTLANDNIKI	67
Db	459	FEFLLANDIDNSNPVQAEQLMNLHYINMIGITIGSEBENEDGYRVAVADVADNLDOI		518
Qy	68	V-DYID--GSD--IDSGDWFTYGNPNPNY--TTPNSQS----	YSL	102
Db	519	ASDYRKAKGAGQSDQAIKHLISLEAMSH--NDAYVYEDTKGAGQLPMDMPHMLAVSL		576
Qy	103	LK--SENSQITOI-----KRYVSGNISRLATKSF-----SAKVTTTSKVI	STON	150
Db	577	LRPIGNRSGVEPLISNLSLNDRBSGKNSGRMANVAFVRAHDEVSIIIGQII--	KNEIN	633
Qy	151	SSRVVINAIDAENFTDELRT-----TKETRENGSYSHK--SSNLSLYVHTW	IPR--	201
Db	634	POST-----GNTFTLDEKKAFEIYNKOMRSGANQTOYINP	SAVALMLTHKDI	687
Qy	202	-----SLKLNMRMEDYNGTVA--Q		221
Db	688	YGDWYTDGQYNAOKSPYYDAITELLKGRIRYAAGQDMKV--NYIGYGN	TNMDA	746

```
Qy      222 SCYKTTGA---DGSSESTRMLAAGSIFPPGNIDGLMLNDIALS-GMAHKS 2688
          ||| ||| : ||| ||| : : : ||| : :
Db      747 SYRYGTGANSADTGTAEATR--NQGMAYIVSNQPALRLTSMILTIMNGAHRN 7966
```

## RESULT 10

T01485  
probable polygalacturonase (EC 3.2.1.15) 1 beta chain F1707.9 - *Arabidopsis thaliana*  
[Species: *Arabidopsis thaliana* (mouse-ear cress)]  
CDate: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 31-Dec-2004  
Accession: T01485  
R.Vyotsetraia, V.S.; Schwartz, J.R.; Toriumi, M.; Yu, G.; Kwan, A.; Oji, O.; Liu, S.; Li, R.;  
rtz, D.; Li, Y.; Palm, C.J.; Shim, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N  
submitted to the EMBL Data Library, June 1998  
A:Description: *Arabidopsis thaliana* chromosome 1 BAC F1707 sequence.

Query Match	6.1%	Score 108;	DB 2;	Length 626;
Best Local Similarity	20.7%	Pred. No. 3.9;		
Matches 65;	Conservative 37;	Mismatches 122;	Indels 90;	Gaps 14

```

QY      26 NNCNFIYANGMNIIGVITINITPTDDBGNFPIDIDVTLNDNIKIVDYIDGS----- 74
DB      172 ESNVVEEGFNSYGF-----GTGAGADFNTYONNVNPPISRTFAYSDDGNGNSQTRKTYT 227
QY      75 -DIDSDGCFYVGNBNEYVTIIPNS--QSYILKSENSOITQIKRYVYCSNTRSRLRTKSPSA 132
DB      228 HEANANGGOSFSTYSGKNGVNETSTSYV--SSNVISGSGFNSYSGSN-----AA 276
QY      133 KYTTTS-GKVISITONSINSRVLINAI-----DATNFTDDELRTTKETREFENOSYT 163
DB      277 NDTFTSYSGDGNVPQNNFPNNYGASGAAVDTFANYRDXANVGDD-----SFS 323
QY      164 SHKSTNSLIYHTWTIIPNSKLQNNRWEVDYNNGMTVAOSCY--YKTGADGSGSESTRMLAA 241
DB      324 SYAKXONSSEKVN-----FVNYGOSFNPBGSEFTFGYGABESKLSFKTYTP 369
QY      242 GSIFPPGNATGLMLNDIALSGMAHKSNNVVDGINGLSTRIIGKFS--WV-----YN 293
DB      370 NSTF-----KDYAKKGVAPAKNVST-----TANTVGDGDKTYNKMIBPGKEFR 413
QY      294 ISGLDRGHAVIDIID 307
DB      414 BSLKSGTVIEMPD 427

```

## RESULT 11

hypothetical protein BB0398 - Lyme disease spirochete  
C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 09-Jul-2004  
C:Accession: E70149  
R:Praser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kervavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugr,  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.  
A:Reference number: A70100; MIMD:198055943; PMID:9403685  
A:Accession: E70149  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA  
A/Residues: 1-343 <KLE>  
A/Cross-references: UNIPROT:O51359; UNIPARC:UP10000057476; GB:AE001145; GB:AE00783; NIT  
A/Experimental source: strain B31  
C/Superfamily: Borrelia burgdorferi hypothetical protein BB0398

Query Match 6.0%; Score 107.5; DB 2; Length 343;  
Best Local Similarity 18.4%; Pred. No. 1.9;  
Matches 44; Conservative 38; Mismatches 80; Indels 77; Gaps 9;

QY 14 VSTFNVILETGLDNCNIYANGLMNIGVINITPTDDEGNFVDIDVTLDNLIKVDYIDG 73  
DB 141 INYKFLIMNGIDILN-----NEBKGNFYMLALSLED---VQDY--- 177  
QY 74 SDDSGSGMFTYTGNNRYNTIP-----NSQSYSLKSNQITQ 112  
DB 178 -----DESYFY---KKFLIPRAHLKIDRDYFNVTIKYFNFPFVYRMGDIQD 229  
QY 113 IKRVSCSNSTRL-----RTKSPSAKYTTSGKVISITONSINSRVVINAIDATNPTD 167  
DB 230 VKRFVLSGNTSKLNTIDKNNFQSWDQKGGK-----SNSINTNSFLTMIIRLGRKK 284  
QY 168 ELRTYKTRFENOSYTSKSTNSLYVHTW-----TIPRSLKQNRWEDYNNGW 219  
DB 285 GIQFARKLLEADSSDDISYLESRGMDHHEWYFVFKRIIVPDD-----PEINNGWTW 335

RESULT 12  
H90053  
hypothetical protein fmb [imported] - Staphylococcus aureus (strain N315)  
C/Species: Staphylococcus aureus  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C/Accession: H90053  
R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hirata, K.  
Lancet 357, 1225-1240, 2001  
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A/Reference number: A89758; PMID:21311952; PMID:11418146  
A/Accession: H90053  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1038 <KUR>  
A/Cross-references: UNIPROT:Q99RD2; UNIPARC:UP100000CAB43; GB:BA000018; PID:G13702453; F  
A/Experimental source: strain N315  
C/Genetic: fmb  
A/Genetic: fmb

Query Match 6.0%; Score 107.5; DB 2; Length 1038;  
Best Local Similarity 22.9%; Pred. No. 8.5;  
Matches 78; Conservative 39; Mismatches 130; Indels 93; Gaps 17;

QY 13 GVSTF-----NVILETG--LDNCNIYANGLMNIGVINITPTDDEGNFVDIDVT 61  
DB 249 GVSTARKVPEIKNGSVVMATGEILGNGNIRYTFNIEHKVEYANLEIMLFDPKTVQS 308  
QY 62 NNDIKIVDYIDSGSDSDGMPYTGNNRYNTIP--SQSYSLKSNQITQIKRYVSCS 120  
DB 309 NGBOKITSKLNGETEKTIPIVY--NPGVNSYNNVNGSLETENKSNKETHI--AYIKPM 365  
QY 121 NTSRLRTKSPSAKYTTSS-----GKVISITONSINSRVVINAIDATNPT 165  
DB 366 NMGOSNTVSTGTLTSSNLAGGPYKRYEYKQDELQOS-----VYANTSDTNKKF 419  
QY 166 DDELRTTKETR-----FENOSYTSKSTNSLYV--HTWTIPRSLKQNRWEDYN--- 214  
DB 420 D---VTKENAGKLSVDNGSYSLMDKLDKTYVIHTGEXLQSSDQVNFRTELYGYPER 475  
QY 215 -----NQM--TWAS--CYKGTGADGSESTRMLAAGSIFPGANTDGLMLNDI--- 259  
DB 476 AYKSYVYVGGYRLTMDGLVLYSKADGNGKN-----GGITQ-----DNDPEYKE 520  
QY 260 -----ALSGMAHKSYNVDTGINGLS-----FTRIIGK 287

DB 521 DTAKGTMGQYDAKQILFEFENDNTPLDIDYHTALDGE 560

RESULT 13  
A86827  
hypothetical protein ydfg [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C/Species: Lactococcus lactis subsp. lactis  
C/Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004  
C/Accession: A86827  
R/Bolotin, A.; Winkler, P.; Mager, S.; Jallou, O.; Malarme, K.; Weisenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp  
A/Reference number: A86825; PMID:21235166; PMID:11337471  
A/Accession: A86827  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1072 <STO>  
A/Cross-references: UNIPROT:Q9CF64; UNIPARC:UP100000D4491; GB:AE005176; PID:G12724625; P  
A/Experimental source: strain IL1403  
C/Genetic: ydfg  
A/Genetic: ydfg

Query Match 6.0%; Score 107; DB 2; Length 1072;  
Best Local Similarity 21.2%; Pred. No. 9.7;  
Matches 65; Conservative 52; Mismatches 136; Indels 54; Gaps 10;

QY 10 DMTGVSTFNVILETGLDNCNIYANGLMNMT--GVINITPTD-----EGNFVDID 57  
DB 610 DSEISTGAFSLSNQTSSEASMSMSINSPLSLTSSSATQNSSEATKYDNN 669  
QY 58 DVTLDNLIKVDYIDSDIDG---SPGMFTYTGNNRYNTIPNSQSYSLKSNQITQIK 114  
DB 670 SSTHSSNTLNSGNSDSDSDSDSSNSLSSPN---LETQITSSKPEVNNISENP 725  
QY 115 RYVSCSN-----SLRTKSPSAKYTTSGKVISITQ--NSINSRVVINAIDATN 163  
DB 726 KKVSSNSVQENSTDMHMTNPKSSISSTSTSSQKESQSNLNTTGGINNPIFPNN 785  
QY 164 FTDELRTTKETREPENOSYTSKSTNSLYVH-----TWITPRSLKQNRWED 212  
DB 786 SSSSENSAASITLTSYNNN--SESSFTGCLYISNDAQRDNGSEIHSLSPSNSNEN----- 838  
QY 213 YNNGWTWASQCYKGTGADGSESTRMLAAGSIFPGYDGLMDNIDALSGMAHKSYNVD 272  
DB 839 -NVSSTQQAILESSKSTNKRSSLSLITNSTHPQNE-----DNQSNSDVYKSNNV 892

QY 273 TGINOLS 279  
DB 893 SLIGQLN 899

RESULT 14  
JH0284  
125k surface antigen M17 precursor - Entamoeba histolytica  
C/Species: Entamoeba histolytica  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C/Accession: JH0284  
R/Edman, U.; Meraz, M.A.; Raueser, S.; Agabian, N.; Meza, I.  
J. Exp. Med. 172, 879-888, 1990  
A/Title: Characterization of an immuno-dominant variable surface antigen from pathogenic  
A/Reference number: JH0284; PMID:90354789; PMID:1696956  
A/Accession: JH0284  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-1114 <EDM>  
A/Cross-references: UNIPROT:Q05352; UNIPARC:UP1000007AD25  
A/Experimental source: strain HMI IMS

Query Match 6.0%; Score 107; DB 2; Length 1114;  
Best Local Similarity 21.8%; Pred. No. 10;  
Matches 77; Conservative 37; Mismatches 125; Indels 114; Gaps 19;

QY 21 LETGLDNCNIYANG--LNMIGVILNITPTDEGNFVIDDV--TLNDNIKIYVIDGSD 75  
DB 187 IQVGIGKCNINPBDQMLANVMNMSNDRMFIDSIGEDGLNTQYIINDTFTKI-----GSP 241  
QY 76 IDG-----SD-----GWFYTNPN-----EYNTIPNSQSYSLKSENSQITQIKRYVSC 119  
DB 242 FGGMIYLRSDTFTNSFYVTFNSVGRAPITINYNTTBEWNSVLRNAPGNVAEIR----- 296  
QY 120 SNTSRLRTKSPSAKVTTTSGKVISITONSINS-----SRVYNALDANFTDDELRTT 172  
DB 297 -----TPGNRLVLTISRNIRSLDQAQYISDFWLKALISINYA----- 332  
QY 173 KETRFENOSYTSH-----KSTNSLYVHTWTIPRSLKLONMWEDYNNGWMTAOSCYKKT 227  
DB 333 --VTLENIPITLNFDPQVDAVAAYVGRWFT-----QNPSDMAAC---V 373  
QY 228 GADGSESTRW--LAAGSIFPPGNY--DGLMLNDIALSGMAKSYNVDTGINQLSFTRI 283  
DB 374 GKDGLINYGWNGPLHEMNHHMQGYLKGHW-----GISNPGERTNNVMTSINYIYLTNI 428  
QY 284 IG---KGFS--WVYNISGLDRGHAVIITDOYGNKRYILFHAGYENSDEPLSSI 332  
DB 429 AGHRNQLSGMNY-----VSDGYSTYKIL--KGBNDQPHLSYV 466

## RESULT 15

B81963  
IGa-specific serine endopeptidase (EC 3.4.21.72) MMA0457 [imported] - Neisseria meningitidis  
C/Species: Neisseria meningitidis  
C/Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
C/Accession: B81963  
R/Parikhil, J.; Achtmann, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel,  
Nature 404, 502-506, 2000  
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A/Reference number: A81775; MUID:20222556; PMID:10761919  
A/Accession: B81963  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1,1449 <PAR>  
A/Cross-references: UNIPROT:Q9TMB4; UNIPARC:UPI00000C4995; GB:AL162753; GB:AL157959; NIT  
A/Experimental source: serogroup A, strain Z2491  
C/Genetics:  
A/Gene: iga2; MMA0457  
C/Superfamily: Iga-specific metalloendopeptidase  
C/Keywords: hydrolase; serine protease

Query Match 6.0%; Score 107; DB 2; Length 1449;

Best Local Similarity 21.1%; Pred. No. 15;  
Matches 71; Conservative 46; Mismatches 126; Indels 94; Gaps 17;

QY 9 SDWTVGTFNVILFTGLDNCNIYANGLMIGVILNITPTDEGNFVIDD-----V 59  
DB 722 SDWTVGLT--NCVEKRTTDD-----KVIALSTKTTDISGN-VSLADHAHLNLGLA 767  
QY 60 TLNDNIKIYVIDGSDIDGSDGWFYTGPNENYNTIPN--SQSYSLKSENSQITQIKRYVS 118  
DB 768 TLNGNLS-----ANGDTRYVSHATONGNLSIYGNAAQT--FNQATL 808  
QY 119 CSNTSRLRTKSPSAKVTTTSGKVISITONS--INSRVYNAIDAFTDDELRTTKETRF 177  
DB 809 NGNTSAGNAPFNLSNNAQNGSLTLSDNAKANVSHSLNG-----NVSADKAVFHF 861  
QY 178 ENQSYTSKHSSTNSLYVH---TWTPRSLKLONMWEDYNNGWMTAOSCYV--KTGADG 231  
DB 862 ENSRFTGQLSGSKDTALHKDSEWTLPSGTEGLNL--NLDNAITTLNSAYRHDAAGQT 918  
QY 232 GS-----ESTRWLAAGSIFPPGNYDGLMLNDIALSGMAKSYNVDTGINQLSFTRI 284  
DB 919 GSVSDTPRRRRRRRL--SVTPPTSVESRF--NLITVNGK-----LNGGTRFRM 964  
QY 285 GKGFWS-----VYNISGLDRGHAVIITDO 308

DB 965 SELFGRSDKLKLAESSEGTYYTLAVNNTGNBPVSLDQ 1001

Search completed: January 30, 2006, 09:49:39  
Job time: 7.82461 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 30, 2006, 09:40:48 ; Search time 26.6268 Seconds

(without alignments)  
8849.971 Million cell updates/sec

Title: US-09-889-874A-22

Perfect score: 1784

Sequence: 1 FTLRSDMSDMVTGFSTFNV.....ILFHAGYNSDPYLSSTIVY 334

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: uniprot\_05.80:.\*  
2: uniprot\_trembl:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1749	98.0	327	09EVR8_XENBV	09EVR8 xenorhabdus
2	145.5	8.2	340	07NAU4_PHOUL	07NAU4 photobacteri
3	136	7.6	334	052884_COXBU	052884 coxiella bu
4	129.5	7.3	1550	04JL63_LACRE	04JL63 lactobacilli
5	126	7.1	892	05HR16_STAEQ	05HR16 staphylococ
6	124.5	7.1	931	09K113_STABP	09K113 staphylococ
7	124	7.0	293	04N914_THBPA	04N914 thelateria p
8	124	7.0	1056	08C072_STARP	08C072 staphylococ
9	124	7.0	1092	070022_STARP	070022 staphylococ
10	122.5	6.9	404	093109_VVIBR	093109 vibrio pomm
11	122.5	6.9	990	097VMO_SULSO	097VMO sulfolobus
12	121	6.8	1765	07V8S5_PROKM	07V8S5 prochlorococ
13	119	6.7	814	052984_ECOLI	052984 escherichia
14	119	6.7	883	083287_ENTRA	083287 enterococcu
15	119	6.7	1361	04L555_STAHJ	04L555 staphylococ
16	118.5	6.6	537	054WU1_DICDI	054WU1 dictyostell
17	118	6.6	773	07UCU4_SHIFL	07UCU4 shigella fl
18	118	6.6	1502	07R8L6_PLAYO	07R8L6 plasmodium
19	118	6.6	1502	07R8L6_PLAYO	07R8L6 plasmodium
20	118	6.6	6761	06L8X8_PLAF7	06L8X8 plasmodium
21	117.5	6.6	627	07BD11_PLAYO	07BD11 plasmodium
22	117.5	6.6	640	APRN_ENTHI	APRN entamoeba h
23	117.5	6.6	3779	05GQB9_9CAUD	05GQB9 bacterioph
24	117	6.6	941	07R1W5_PLAYO	07R1W5 plasmodium
25	116.5	6.5	1054	09LAB9_GRICK	09LAB9 grammycete
26	116.5	6.5	1092	08BUS1_MYCPE	08BUS1 mycoplasma
27	115.5	6.5	861	08PVN5_METMA	08PVN5 methanosa
28	115.5	6.5	962	05S854_DICDI	05S854 dictyostell
29	115.5	6.5	1670	023901_DICDI	023901 dictyostell
30	115.5	6.5	1671	086985_DICDI	086985 dictyostell
31	115.5	6.5	2835	08G9Q2_LEUME	08G9Q2 leuconostoc

32	115.5	6.5	3623	07R069_PLAYO	07R069 plasmodium
33	115	6.4	672	06FK03_CANGA	06FK03 candida gla
34	115	6.4	2001	06KCM6_ECOLI	06KCM6 escherichia
35	115	6.4	2387	08FFP8_ECOLI	08FFP8 escherichia
36	114.5	6.4	456	06CUX7_KLULA	06CUX7 kluyveromyc
37	114.5	6.4	1392	04QNU0_HAB18	04QNU0 haemophilus
38	114.5	6.4	3078	08IKH4_PLAF7	08IKH4 plasmodium
39	114	6.4	5251	08IKH4_PLAF7	08IKH4 plasmodium
40	113.5	6.4	878	06B196_DEBHA	06B196 debaryomyce
41	113.5	6.4	1054	09LAF1_GRICK	09LAF1 rickettsia
42	113	6.3	412	054X20_DICDI	054X20 dictyostell
43	113	6.3	584	09K113_MYCMA	09K113 mycoplasma
44	113	6.3	2110	08IBF2_PLAF7	08IBF2 plasmodium
45	112.5	6.3	608	04KH09_PLACH	04KH09 plasmodium

## ALIGNMENTS

## RESULT 1

09EVR8\_XENBV  
ID 09EVR8\_XENBV PRELIMINARY; PRT; 327 AA.  
AC 09EVR8;  
DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)  
DE Nematocidal protein.  
GN Name=xmpl;  
OS Xenorhabdus bovienii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Xenorhabdus.  
OX NCBI\_Taxid=40576;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=173;  
RA Morgan J.A.W., Ellis D., Jarrett P.;  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AJ296651, CAC19492.1, -; Genomic DNA.  
SQ SEQUENCE 327 AA; 36884 MW; 8417B96FB70D60D CRC64;

QY	8	MSDWTGVSTFNVLTETGDNCTNYANGLMVGIIINTPTDDEGNFVIDDVTIANIKI	67	Query Match	98.0%; Score 1749; DB 2; Length 327;
Db	1	MSDWTGVSTFNVLTETGDNCTNYANGLMVGIIINTPTDDEGNFVIDDVTIANIKI	60	Best Local Similarity	100.0%; Pred. No. 1.7e-116;
QY	68	VDYIDSGSDIDSGDWGYTGNPNRYNTPNSQSYSLKSENSQITQIKRYVCSNTRSLRT	127	Matches 327; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	61	VDYIDSGSDIDSGDWGYTGNPNRYNTPNSQSYSLKSENSQITQIKRYVCSNTRSLRT	120		
QY	128	KSPSAKVTTTSGKVISITONSINSSRVVINAIDATFTDELARTTKETREPNQSYTSHKS	187		
Db	121	KSPSAKVTTTSGKVISITONSINSSRVVINAIDATFTDELARTTKETREPNQSYTSHKS	180		
QY	188	STNSLVVHTWTPRSLKQNMREDDYNGWTMAQSCYTKTGADGSGSTRMLAAGSIFPP	247		
Db	181	STNSLVVHTWTPRSLKQNMREDDYNGWTMAQSCYTKTGADGSGSTRMLAAGSIFPP	240		
QY	248	GNDVGLMLNDIALSGMAHKSYYVDGTGINOLSTRTIIGKGFSSVWVNSIGLDRGHAVIIID	307		
Db	241	GNDVGLMLNDIALSGMAHKSYYVDGTGINOLSTRTIIGKGFSSVWVNSIGLDRGHAVIIID	300		
QY	308	QYGNKRTIIFHAGYNSDPYLSSTIVY	334		
Db	301	QYGNKRTIIFHAGYNSDPYLSSTIVY	327		
RESULT 2					
Q7NAU4_PHOUL					
ID	Q7NAU4_PHOUL	PRELIMINARY;	PRT;	340	AA.
AC	Q7NAU4;				

DT 01-MAR-2004 (TReMBLrel. 26, Created)  
 DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
 DE Unknown protein.  
 GN OrderedLocustNames=plu2221;  
 OS Photobacterium lumenescens (subsp. laumondii).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Photobacterium.  
 OX NCBI\_TaxID=141679;  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=TT01;  
 RX MEDLINE=D295627; PubMed=14528314; DOI=10.1038/nbr886;  
 RA Duchaud E., Rustiock C., Frangeul L., Buchrieser C., Givaudan A.,  
 RA Taourit S., Boco S., Bouraux-Eude C., Chandler M., Charles J.-F.,  
 RA Daba B., Desobe R., Derzelle S., Freysinet G., Gaudinault S.,  
 RA Meligne C., Lanois A., Powell K., Siguler P., Vincent R., Wingate V.,  
 RA Zouine M., Glaeser P., Boemare N., Danchin A., Kunst F.,  
 RT "The genome sequence of the entomopathogenic bacterium Photobacterium  
 lumenescens";  
 RL Nat. Biotechnol. 21:1307-1313(2003).  
 DR EMBL; BX571866; CAE14514.1; -; Genomic\_DNA.  
 DR PhotoLib; plu2221; -;  
 KW Complete proteome.  
 SQ SEQUENCE 340 AA; 38470 MW; 818864919F82175E CRC64;

Query Match 8.2%; Score 145.5; DB 2; Length 340;  
 Best Local Similarity 22.9%; Pred. No. 0.033;  
 Matches 74; Conservative 59; Mismatches 137; Indels 53; Gaps 15;

QY 29 NIYANGLMIGVINIPTPTD-DEGNFVIDDVTINDNKK-----IYDIIDGSDIDGSD 80  
 DB 20 NFYANGRHQCWVKISVLKQEKYKGDWVCL--ALSDAKRSIQVAAALSDSLIYQOLKMS 76  
 QY 81 GWFYGNPNEX-----NTIPNSQSYS-----ILKSENSQ--ITQIKRVYSC-SNT 122  
 DB 77 GMTTDDANKKDYDLGLANGVYADIFTEBPVIRAGDCCCTNENYQNSVKSPEIICYVSS 136  
 QY 123 SRLRTKSSAKVT--TISGKVIISTGNSIN-----SSRVINAIIDANTDELATTKETR 176  
 DB 137 NRTSTEYIMAKWTFEDINGKRTLTNNMSVGDEVEDSKILLALIPYAINAQULHSTNITL 196  
 QY 177 FENOSYTHSKSSTN--SLYVHTWTIPRLKLONNRWMEDYNNGWTAQSCYYKTGADGSE 234  
 DB 197 FDKTEEPF-KSDTHHQITINLRWTLPIHLRLBEGDRVNNIYVLGKS-----SS 245  
 QY 235 STRWLAAGSIFPPGNYDGLMLDNDIALSGMAHKSYNVDGTGINQSFTRIICKGFSWVYNI 294  
 DB 246 DDFLTRARVFPKRGF-SYVNAARNDMSSGCGVMDYSYDVTVPQTOLAABVLAVHTGCGSWTGY 304  
 QY 295 SGLDRGHAIVIIIDYG--NKTRI 315  
 DB 305 --VDGYHDVTIIDNYGCOHKFRI 325

RESULT 3  
 052884 COXBU  
 ID 052884 COXBU PRELIMINARY; PRT; 334 AA.  
 AC 052884;  
 DT 01-JUN-1998 (TReMBLrel. 06, Created)  
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Hypothetical protein ort 334b.  
 GN Name=ort 334b;  
 OS Coxiella burnetii.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
 OC Coxiellaceae; Coxiella.  
 OX NCBI\_TaxID=777;  
 RN NUCLEOTIDE SEQUENCE.  
 RP Laurentschlaeger S., Jaeger C., Willems H., Baljer G.,  
 RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; Y15898; CAAY5845.1; -; Genomic\_DNA.  
 KW Hypothetical protein; Plasmid.  
 SQ SEQUENCE 334 AA; 39048 MW; 6BDC7BE729851DDC CRC64;

Query Match 7.6%; Score 136; DB 2; Length 334;  
 Best Local Similarity 21.3%; Pred. No. 0.15;  
 Matches 77; Conservative 49; Mismatches 155; Indels 80; Gaps 17;

QY 11 WTGVSTFNVILF-----GLDNCNIYANGLMIGVINIPTPTDEGNFVIDDVTINDNKK 66  
 DB 3 WENPFRSIVKLAIPGHPTNQARIYANGRNQVIRVLRLDREGNPLKLTIEDLREKCS 62  
 QY 67 IYDIIDGSDID-----GSDGWFPYGNPNEX-----NTIPNSQSYS--LKSENSQ 109  
 DB 63 LIYQYREEBISYKSGENSESANISSQNWYGTGYKNDYQCNFPLVP-----YSLNTNDSDDK 118  
 QY 110 ITQIKRVYSCNTRSLRTSFPKAVTTTSGKVIISTGNSINSGRVYINAIIDANTPTDEL 169  
 DB 119 VSYIDYTL-----KSGERTGKHIGLCINIGENYKLSGKFKGGVYETIAPKKYTDNI 174  
 QY 170 RTTKETREFNQSYTHSKSSTNSLYVHTWTIPRLKLONNRWMED-----YNNGWTA 220  
 DB 175 TFRRESK--KIGTTNWVGAQGYNVEKTYLTIKQSFRIKKGIDINNVFIRYGDWK 232  
 QY 221 QSCYKTKGADGSGSESTRWLAAGSIFPPGNYDGLMLDND-IALSGMAHKSYN----- 270  
 DB 233 FGFFLDPSGD-----KRTGQHLVPIEFS--LRDVKVFFSGKYHVQNVFFYEDAMC 281  
 QY 271 --VDGTINQSLFTRIICKGFSWYNNISGLDRG-----AVIIDIQYEN--KIRLFLPAGY 321  
 DB 282 FFWVTGV--FVPIIQQ--WTY--GNEKCNEMKCFYFYDQYNGEKFTRICDDGW 331  
 QY 322 E 322  
 DB 332 E 332

RESULT 4  
 04JUG3 LACRE  
 ID 04JUG3 LACRE PRELIMINARY; PRT; 1550 AA.  
 AC 04JUG3;  
 DT 13-SEP-2005 (TReMBLrel. 31, Created)  
 DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORNames=lr1612;  
 OS Lactobacillus reuteri.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillaceae;  
 OX NCBI\_TaxID=1598;  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=ATCC 55730;  
 RA Wall T., Bach K., Roos S., Jonsson H.,  
 RT "Bioinformatic identification of the secretome of Lactobacillus  
 reuteri ATCC 55730."  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; DQ074912; AA186887.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1550 AA; 172371 MW; 9A7052287A9014D7 CRC64;

Query Match 7.3%; Score 129.5; DB 2; Length 1550;  
 Best Local Similarity 23.7%; Pred. No. 2.8;  
 Matches 76; Conservative 55; Mismatches 139; Indels 51; Gaps 16;

QY 39 GVIINPTPTDEGNFVDI-DDVTIND-----NIKIYDIIDGSDIDGSDGW 82  
 DB 69 GIINPTTPASAEKNSVCGSGSTVNDQTSPOVDSNIYGVNVVMDVYENNSDIYNLTG- 127  
 QY 83 FYTGPNPNENYNTIPNSQSYSLKSENSQITQIKRVYSCNTRSLRTKSSAKVTTTSG--- 139  
 DB 128 -YHGDTRHIVVPLINDFNDAIGIIGTTSVG--ISSITLTKNLVQSIAFAISKISGYN 184



```

QY      140  -KYISTON---SINSRVINA-----IDACFTGDEL-----RTTETRENOSSYSHK 186
Db      185  NKVIA-TDNDMSVFPASSGVLNADLSNIDTSITNMKYLPAVNGKLTGVLISEMDTGHV 243
QY      187  S8TNSLVHTWTTPRSLKLONRWEDYNNGWMTWAOSCYTKGADGSESTEWILA6IFP 246
Db      244  SDMA5HFYE5ALQ5AGNIGKMKNTENVTYSMMNMFPESDVSTKRTIGHNNTSKVVDMGTF- 302
QY      247  PCNYDGLM--LDN-DIALSGMAHKSYNDVTGINOLSPFRIIGKPSW--VYNISGLDRGHA 302
Db      303  -HSYKCEWGLDSNWDVSNVKNMVEFWSSDDRGLSEINPGNITK---WKIKGANLVR--- 355
QY      303  VIIIDQGNKYRLFLPAGYEN 323
Db      356  MEF5FDRSTSPYRVV--ASNEN 374

```

```

SRH16_STAEO
5 OSHR16_STAEO PRELIMINARY; PRT; 892 AA.
AC OSHR16;
DT 10-MAY-2005 (TReMBLrel. 30, Created)
DT 10-MAY-2005 (TReMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TReMBLrel. 30, Last annotation update)
DE Sarg protein.
OS OrderedlocusNames=SERP0207;
OS Staphylococcus epidermidis (strain ATCC 35964 / RP62a).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
NCBI_TaxID=176279;
RN
RP NUCLEOTIDE SEQUENCE [Large scale genomic DNA].
RX PubMed=15774885; DOI=10.1122/DB.187.7.2426-2438.2005;
GI|18.9, Fouts D.E., Archer G.L., Mongodin E.F., DeBoy R.T.,
RA Ravel J., Paulsen I.T., Kolonay J.F., Brinkac L.M., Beaman M.J.,
RA Dodson R.J., Daugherty S.C., Madupu R., Angiuoli S.V., Durkin A.S.,
RA Haft D.H., Vamathevan J.J., Khouri H., Uetcheback T.R., Lee C.,
RA Dmitriev G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H.,
RA Hance I.R., Nelson K.E., Fraser C.M.,
RT "Insights on evolution of virulence and resistance from the complete
RT genome analysis of an early methicillin-resistant Staphylococcus
RT aureus strain and a biofilm-producing methicillin-resistant
RT Staphylococcus epidermidis strain."
RL J. Bacteriol. 187:2426-2438(2005).
DR EMBL: CP000029; AAM53571.1; -, Genomic_DNA.
DR SRR; OSHR16; 237-557.
DR TIGR; SERP0207; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KM Cell wall; Complete proteome.
SQ
SEQUENCE 892 AA; 98746 MW; EC35C932B58EF497 CRC64;

Query Match 7.1%; Score 126; DB 2; Length 892;
Best local similarity 21.0%; Pred. No. 2.5;
Matches 78; Conservative 51; Mismatches 141; Indels 102; Gaps 16

QY 9 SDWTGSTNVLELTGLDNCNLYANGLMWG--VIINTPDDSGNFVDDPYT----- 60
DB 398 SMTNIDTKNTHVEQ-----TIYNPLKRSAKETNVNLSGNGDGSTI-IDDSITIKVYK 451
QY 61 -----LNDNIKIVDITDSDIDGSDGMPYTGPNPEVNT-----IPNS 97
DB 452 VEGDNQLPDSNRIRYDYSSEYEDVT-HDDVAQGNNDNVINNGNIDSPYIKVISKDPNK 510
QY 98 QSSLSLAKSENSQTOIKRIVYCSNTRSLRTSFSKAV--TTSG-----KVSI-- 144
DB 511 DDVTTTQQATVMTQTTINEX-----TGEFFTASVDYNTIAFSTSSGQGGDLPPEKTYVIGD 565

```

Qy 143 -----TQNSINSSVVVNAIDANFNPDELRTTKERFRSNOGYTHSKSTSL Y 193  
Db 566 YWEDVDKDGIGINTNDNEKPLSNVLTLLTPDG--TSKSVRDEBKGKYPFDGKGLTY 622  
Qy 194 VHTWTIPRLAKLQNRMEDY-----NNGWTMAQS CYRTKGDGSEST 236  
Db 623 KITFFRP-----EGYPTPLHGSTGNALDSEGNVAVTTINGDDMTIDSGPQTR 672  
Qy 237 RMLAAGSIFPPGNVYDGLMFLDIALSGNAHKSIVYDTGINKLSFTRIIIGKFSWYINISG 296  
Db 673 KYSLSGNYVWYDITNKGIGIQGDDDEKGISGV-KVTLKDENGNIISITTTDENG--KIQPDN 727  
Qy 297 LDRGHAVITIDQ 308  
Db 728 LNSGNYIVAFDK 739

ID	Q9KIL3	STRAEP	PRELIMINARY;	PRT;	931 AA.
AC	Q9KIL3				
DT	01-OCT-2000	(TRMBLrel. 15, Created)			
DT	01-OCT-2000	(TRMBLrel. 15, Last sequence update)			
DT	01-OCT-2003	(TRMBLrel. 25, Last annotation update)			
DE	Fibrinogen-binding protein SdG.				
GN	Name=sdg;				
OS	Staphylococcus epidermidis				
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.				
OX	NCBI_TaxID=1282;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=K28;				
RX	MEDLINE=20340957; PubMed=10878118;				
RA	McCrea K.W., Hartford O., Davis S., Ni Eidhin D., Lina G.,				
RA	Speziale P., Foster T.J., Hook M.;				
RT	"The serine-aspartate repeat (Sdr) protein family in Staphylococcus				
RT	epidermidis";				
RL	Microbiology 146:1535-1546(2000).				
DR	EMBL; AF245042; AAF72510.1; -; Genomic_DNA.				
DR	PDB; 1R17; X-ray; A/B=274-598.				
DR	PDB; 1R19; X-ray; A/B/C/D=274-598.				
DR	GO; GO:0009986; C:cell surface; IEA.				
DR	GO; GO:0005618; C:cell wall; IEA.				
DR	GO; GO:0016020; C:membrane; IEA.				
DR	Interpro; IPR008454; Cna B.				
DR	Interpro; IPR005877; Gpos_YsIRK.				
DR	Interpro; IPR001899; Gram_pos_anchor.				
DR	Pfam; PF05738; Cna_B; 2.				
DR	Pfam; PF00746; Gram_pos_anchor; 1.				
DR	Pfam; PF04650; YsIRK_signal; 1.				
DR	TIGRFAMs; TIGR01167; LPXTG_anchor; 1.				
DR	TIGRFAMs; TIGR01168; YsIRK_signal; 1.				
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.				
KW	Cell wall.				
SO	SEQUENCE	931 AA; 102955 MW; 5918657D97027116 CRC64;			
Query Match		7.1%; Score 126; DB 2; Length 931;			
Best Local Similarity		21.0%; Pred. No. 2.7;			
Matches	78; Conservative 51; Mismatches 141; Indels 102; Gaps 16;				
QY	9	SDMWGVSTFNVVILETGLDNCNIYANGLMTG--VIINITPRDDSGVNDIDDT-----	60		
DB	437	SMFNTIDTKNHYEQ-----TIIYNPLRYSAKETNVNISGNGDGSTI-IDDSITIKYK 490			
QY	61	-----LNDNIKIVDYIDGSDIDSDGMFYTGNDENEYNT-----IPNS 97			
DB	491	VGDNQNL.PDSNRIDYSEYEDVT-NDQYALGNNDVNIINGNIDSPYIIKIVISKYPENK 549			
QY	98	QSVSLKSEKSGQITQIRGVYSCNTSLRTKRSPEAKY--LTTSG-----KVISF-- 144			
DB	550	DDVYITQIVYMTQTTINEY-----TGEFRITASYDNTITAFSTSSGGGGDLPEPKTYKGD 604			
QY	145	-----TONSINSSRVVINAIDTFTDELRTTKETRPENOSYTSKKSSTNSL-Y 193			

Db 605 YWEDVDKQGIQNTNNEKPLSNVLVTLYPDG--TKSVRTTBDEGKYQFDGLKNGLTY 661  
 QY 194 VHTWTTIPRSJLQNWREMDY-----NNGWTWQSCYYKYGADGSEST 236  
 Db 662 KITFERP-----EGYTPPLKHSCTNPALPDBEGSNWVTTINGDDMTIDSGFYQTP 711  
 QY 237 RMTLAAGSIPPGNYDGLMLDNDIALSGMAHRSYNYDTGINSFRTLLIGKGSFWVYNISG 296  
 Db 712 KYSLGNVWYDINKDGIQGDDEKGISGV--KYLKDBNGNIISTTTDENG--KYQPDN 766  
 QY 297 LDRGHAVIITDQ 308  
 Db 767 LNSGNVIYHFDK 778

ID	QAN914	THEPA	PRELIMINARY;	PRT;	293 AA.
AC	QAN914	THEPA	PRELIMINARY;		
DT	13-SEP-2005	(TREMBLrel. 31, Created)			
DT	13-SEP-2005	(TREMBLrel. 31, Last sequence update)			
DT	13-SEP-2005	(TREMBLrel. 31, Last annotation update)			
DE	Hypothetical protein.				
GN	ORFNames=TP01_0300;				
OS	Thelateria parva.				
OC	Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;				
OC	Theileria.				
OX	NCBI_TaxID=5875;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=Muga.				
RA	Gardner M.J., Bishop R., Shah T., de Villiers E.P., Carlson J.M.,				
RA	Hall N., Ren Q., Paulsen I.T., Pain A., Berriman M., Wilson R.J.M.,				
RA	Sato S., Ralph S.A., Mann D.J., Xiong Z., Shallom S.J., Weidman J.,				
RA	Jiang L., Lynn Y., Weaver B., Shoebli A., Masawa D., Crabtree J.,				
RA	Mortman J.R., Haas B., Anguilo S.V., Creasy T.H., Lu C., Suh B.,				
RA	Silva J.C., Utterback T.R., Feldblyum T.V., Petrea M., Allen J.,				
RA	Taracha E.L.N., Salzberg S.L., White O.R., Fitzhugh H.A., Morzaria S.,				
RA	Venter J.C., Fraser C.M., Nene V.				
RT	"Genome sequence of Theileria parva, a bovine pathogen that transforms				
RT	lymphocytes."				
RL	Science 309:134-137(2005).				
RN	[2]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=Muga.				
RA	Gardner M., Bishop R., Shah T., de Villiers E., Carlson J.M., Hall N.,				
RA	Ren Q., Paulsen I.T., Pain A., Berriman M., Wilson R.J.M., Sato S.,				
RA	Ralph S.A., Mann D.J., Xiong Z., Shallom S.J., Weidman J., Jiang L.,				
RA	Lynn Y., Weaver B., Shoebli A., Masawa D., Crabtree J., Mortman J.R.,				
RA	Haas B., Anguilo S., Creasy T.H., Lu C., Suh B., Silva J.C.,				
RA	Utterback T., Feldblyum T., Petrea M., Allen J., Taracha E.L.,				
RA	Salzberg S.L., White O., Fitzhugh H.A., Morzaria S., Venter J.C.,				
RA	Fraser C.M., Nene V.				
RL	Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.				
CC	-!- CAUTION: The sequence shown here is derived from an				
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is				
CC	preliminary data.				
DR	EMBL; AAKG0100001; EAM33544.1; -; Genomic_DNA.				
KW	Hypothetical protein.				
SC	SEQUENCE 293 AA; 31818 MW; 12CF9BC13DC2DF53 CRC64;				
Query Match	7.0%; Score 124.5; DB 2; Length 293;				
Best Local Similarity	22.3%; Pred. No. 0.85;				
Matches	61; Conservative 43; Mismatches 95; Indels 75; Gaps 12				
Qy	27 NCNIYANGLMIGVIINITPTDDEGNFVDIDVTILNDIKIYVDYDGSDDGSMFYTG 86				
Db	63 NCKLIPNKTN-----LNICKIKSFG-FINPSNHTVNSVNSVNTSVNSVNTSV 109				
Qy	87 NPENENVTIPNGSGSYSLKSENSQIQIKRYVCSNTRLRKTSFPAKYTTTSGKVIISTQ 146				
Db	110 -VNSVNTSVNSVNTSVNSVNTSVNSVNTSVNSVNTSVNSVNTSVNSVNTSV 168				

```

Qy 147 NSINSSRVVIAIAINATNTDDELRTKTRFENQSYSTSHKSSSTSLVYVTWTIIPRLKQ 206
Db 169 NSVNTS-----VNSVN-----TSVNSSVNT-----SVNTL 193

Qy 207 NMRREDYNNGMTWAQSCYYK-TGADGSGSESTWLAGSIFPEGNYDG-----LMLDN 257
Db 194 GNDMSDTR-----VVDLCVTKVSGSGDGSGGCSFRRKRV-PLGAGANGSGRGQDYIEC 248

Qy 258 DIALSGM-----AHSYNDVTGINQLSFTRIIGK 287
Db 249 NDSVSDLRMFKANKIKYKQNGNN-----GKG 274

```

ID	Q8C072	STAE	PRELIMINARY;	PRT;	1056 AA.
AC	Q8C072	STAE	PRELIMINARY;	PRT;	1056 AA.
DT	01-MAR-2003	(TrEMBLrel. 23, Created)			
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)			
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)			
DE	Sec-Aap rich fibrinogen-binding, bone sialoprotein-binding protein.				
GN	Ordered locus names=SE0331;				
OS	Staphylococcus epidermidis				
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.				
OX	NCBI_TaxID=1282;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].				
RC	STRAIN=ATCC 12228;				
RX	PubMed12950922; DOI=10.1046/j.1365-2958.2003.03671.x;				
RA	Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,				
RA	Qin Z.-Q., Mao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,				
RA	Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.,				
RT	"Genome-based analysis of virulence genes in a non-biofilm-forming				
RT	Staphylococcus epidermidis strain (ATCC 12228)."				
RL	Mol. Microbiol. 49:1577-1593 (2003).				
DR	EMBL; AE016745; AAC03928.1; Genomic_DNA.				
DR	HSPF; O53653; I067.				
DR	SMR; Q8C072; 278-598.				
DR	GO; GO:0009986; C:cell surface; IEA.				
DR	GO; GO:0005618; C:cell wall; IEA.				
DR	GO; GO:0016020; C:membrane; IEA.				
DR	InterPro; IPR008454; Cna B.				
DR	InterPro; IPR005877; GpoS_YSTRK.				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR003167; tep_3.				
DR	Pfam; PF00738; Cna B_2.				
DR	Pfam; PF00746; Gram_pos_anchor; 1.				
DR	Pfam; PF02412; TSP 3; 1.				
DR	Pfam; PF04650; YSTRK_signal; 1.				
DR	TIGRFAMs; TIGR01167; lpxTG_anchor; 1.				
DR	TIGRFAMs; TIGR01168; YSTRK_signal; 1.				
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.				
KW	Cell wall; Complete proteome.				
SQ	SEQUENCE 1056 AA; 115728 mw; 015869A9E5CA2723 CRC64;				
Query Match	7.0%; Score 124; DB 2; Length 1056;				
Best Local Similarity	21.0%; Pred. No.4.3; Indels 102; Gaps 16				
Matches	78; Conservative 50; Mismatches 142;				
QY	9 SDWTGVSTFNVILETGLDNCNIYANGLMIG-VIINITPTDDEGNFVYDIDVT-----60				
DB	439 SMFTNIDTKRHTVEG-----TIYINPLAYSANETVNVISGNGBGST-IDDSTIIKVKYK 492				
QY	61 -----LNDNKIYDIYIDGSDIDGSGDFYIGNPREYNT-----IPNS 97				
DB	493 VGDNONLPDSNRKIDYSEYEDVT-NDDYAQLGNNDVDVINEGNIDSPYIIKIVISKYDPNK 551				
QY	98 QSYSLKSENSQITQIKRYVSCSNSTRKTSFSAKV--TTTSG-----KVYSI-- 144				
DB	552 DDYTTIQCTVIMQTTINEX-----TGEPRTISYDNTIAFSISSGGGGQDLPPEKTYKIGD 606				

```

Qy 145 -----TONGSRRVAINAIDATNFTDELRTTKETRFENQSYTSHKSTNSL-Y 193
Db 607 YWEDVDKQIGQNTNDNEKRLSNVLVLTYPDG---TSKSVRTDEBQKQPDGLKNGLT 663
Qy 194 VHTWTIPRSIKLQNMWEDY-----NNGWTAAQSCYRTGADGSEST 236
Db 664 KIFETP-----EGYPTPLKHSCTNPALDSEGNVWVITNGDDMTIDSGFYQTP 713
Qy 237 RMLAAGSIFPPGVYDGLMNDIALSGMAKSNVVDGINQLSTFRIIGGFSMVYVNI 296
Db 714 KYSLGNTVWYDTNKDGIQGDDEKISGV--KVTLKDENGNIISTTTDENG---KIQFDN 768
Qy 297 LDRGHAVIDIQ 308
Db 769 LNSGNYIVHFDK 780

```

## RESULT 9

```

O70022 STAAP PRELIMINARY; PRT; 1092 AA.
ID 070022 STAAP PRELIMINARY; PRT; 1092 AA.
AC 070022 STAAP PRELIMINARY; PRT; 1092 AA.
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Fibrinogen-binding protein precursor.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=HB;
RC MEDLINE=9826511; PubMed=9596732;
RA Nilsson M., Frydberg L., Flock J.I., Pel L., Lindberg M., Guss B.;
RT "A fibrinogen-binding protein of Staphylococcus epidermidis.";
RL Infect. Immun. 66:2666-2673(1998).
DR EMBL; Y17116; CAA76638.1; -; Genomic_DNA.
DR PIR; T30214; T30214.
DR HSSP; Q53653; 1N67.
DR SMR; 070022; 278-598.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR005877; GpoS_YsIRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B_2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01167; LpxTG_anchor; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
DR PROSITE; PSS0847; GRAM_POS_ANCHORING; 1.
KW Cell wall; signal.
FT STGNL 51 potential.
FT CHAIN 1 1092 fibrinogen-binding protein.
SQ SEQUENCE 1092 AA; 119292 MW; 6542BC39AAD8984 CRC64;

```

Query Match 7.0%; Score 124; DB 2; Length 1092;

Best Local Similarity 21.0%; Pred. No. 4.5;

Matches 78; Conservative 50; Mismatches 142; Indels 102; Gaps 16;

```

Qy 9 SDWTGVSFVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT----- 60
Db 439 SMTNTIDTKNHYEQ-----TIYNPLRYAKETNVNISGNGDEGSI-IDDSTIIIVYK 492
Qy 61 -----LNDNIKIVDYIDGSDIDSGMFWYGNPNENYNT-----IPNS 97
Db 493 VGDNONLPDSNRRIYDYEVEDVT-NDQYAOAGNNNDVINFGNIDSPYIIKVISKVPNK 551
Qy 98 QASISLKSNSQITQIRKYVSCSTSLKRTKSPSAKV--TTTSG-----KVISI-- 144
Db 552 DDVTTTQOTVYMTTINYE-----TGEFRYASYNNTIAFTSSQGGQDLPPEKTYKIGD 606
Qy 145 -----TONGSRRVAINAIDATNFTDELRTTKETRFENQSYTSHKSTNSL-Y 193

```

```

Db 607 YWEDVDKQIGQNTNDNEKRLSNVLVLTYPDG---TSKSVRTDEBQKQPDGLKNGLT 663
Qy 194 VHTWTIPRSIKLQNMWEDY-----NNGWTAAQSCYRTGADGSEST 236
Db 664 KIFETP-----EGYPTPLKHSCTNPALDSEGNVWVITNGDDMTIDSGFYQTP 713
Qy 237 RMLAAGSIFPPGVYDGLMNDIALSGMAKSNVVDGINQLSTFRIIGGFSMVYVNI 296
Db 714 KYSLGNTVWYDTNKDGIQGDDEKISGV--KVTLKDENGNIISTTTDENG---KIQFDN 768
Qy 297 LDRGHAVIDIQ 308
Db 769 LNSGNYIVHFDK 780

```

## RESULT 10

```

O931J9 GVIBR PRELIMINARY; PRT; 404 AA.
ID O931J9 GVIBR PRELIMINARY; PRT; 404 AA.
AC O931J9 GVIBR PRELIMINARY; PRT; 404 AA.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
OS Vibrio pomeroyensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OX Vibrionaceae; Vibrio.
OX NCBI_TaxID=161725;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=CH-291;
RC PubMed=14680703; DOI=10.1016/S0378-1097(03)00843-7;
RA Torres J., Appel B., Lewin A.;
RT "Cloning and molecular characterization of a unique hemolysin gene of
RT Vibrio pomeroyensis sp. nov.: development of a DNA probe for the
RT detection of the hemolysin gene and its use in identification of
RT related Vibrio spp. from the Baltic Sea.";
RL FEMS Microbiol. Lett. 229:223-229(2003).
DR EMBL; AJ314791; CAC40977.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 404 AA; 45760 MW; 0EB685219A4AC331 CRC64;

```

Query Match 6.9%; Score 122.5; DB 2; Length 404;

Best Local Similarity 21.1%; Pred. No. 1.7; 127; Indels 131; Gaps 20;

Matches 85; Conservative 59; Mismatches 142; Indels 131; Gaps 20;

```

Qy 14 VSTFVILETGLDNCNIYANGLMIG--VIINITPTDDEGNFVDIDVT----- 60
Db 24 IKDPSISL--SGSDRGQIFSGNNGNIBFVN-----VDLBGEIEIWKILIRQDK 72
Qy 74 SDDSDGWFY--TGNPNENYNTI-----PNSQSYSLKSR-----N 107
Db 73 KDIO-ELGWLVRPLQSDSEYYSIHGKYALPNAFIRTTQGNPAPQSRRLSKADQSIQVN 131
Qy 108 SQTQIKRY-----VSCSNTRSL-----RKFSKAYTTTSGV--ISTONGSINS 151
Db 132 SKALQSIENYETPLPTVTHSQTLVYSDVQVTDICAEVTLKSGVYDSCSEVPTTV 191
Qy 152 SRVAINAIDATNFTDELRTTKETRFEN--QSYTSHKS--STNSL--YVHTWTIPRSIKL 205
Db 192 SSTTPVIEPNQSSSDLDADADIEEVNMQHSSSLKALAPALABEFIAIYIDSPDEE 251
Qy 206 QNRWEDYNNGTWAAQSCYRTGADGSESTRLAAGSIFPPGVYDGLMND--IALS 262
Db 252 PLIIVDEYKVPAGEKTAAY-----NVDGINQLSTFRIIGGFSV--WIDNETYHPLP 283
Qy 263 GMAHKSQ-----NVDGINQLSTFRIIGGFSV--WIDNETYHPLP 286
Db 284 NLNNFRLLTGLKLELNLKSVSSKINTLT-----RGFTIILNVVTTIPNMQSTYDTTW 337
Qy 297 LDRGHAVIDIQ 308
Db 338 VVDG---VYNNTYGGCYFPPERNFLQANSTNCKLKSSEIL 376

```

RESULT 11  
 Q97VMO\_SULSO PRELIMINARY; PRT; 990 AA.  
 AC Q97VMO;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocustNames=SSO2602;  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 NCBI\_TaxID=2287;  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=1142726; DOI=10.1073/pnas.141222098;  
 RA Snie Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aweyer M.J., Chan-Weiner C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moors A., Eiruseo G., Fletcher C., Gordon P.M.K.,  
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Regan M.A., Senses C.W., Van der Oost J.,  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 DR EMBL; AE006856; AAA42724.1; -; Genomic\_DNA.  
 DR PIR; E90433; E90433.  
 KM Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 990 AA; 111429 MW; DBE2517A7FBA596A CRC64;

Query Match 6.9%; Score 122.5; DB 2; Length 990;  
 Best Local Similarity 20.8%; Pred. No. 5.1;  
 Matches 64; Conservative 45; Mismatches 99; Indels 99; Gaps 13;  
 QY 36 NMIVIIINPTDEGNFVIDD-----VTLNDIKIVYIDGSDIDGSDGW 82  
 DB 728 NLVVISLKGPTSN--NVIYNGPVSLKHYVLYQYINLISNIPKALINSTLVTPSGW 785  
 QY 83 FYTGNP---NEYTIPNSQSYSL-----KSENSQITQIKRVYSCSNTSRRTK 128  
 DB 786 YNAVTPISFIMTYIISNTRYIILSLIPFVTVRSITVAVTKL----- 834  
 QY 129 SFSAKVTTSGKVTISITONSINSKRVYVNAIDATNFDDERTKTRFEXQSTSHKS 188  
 DB 835 -----VTVNEPILITIANRINTNTSEIWPV-----CQTLIPKYVVISNBERIFYNTS 882  
 QY 189 TNSLYVHWTLPRSLKQNMWRMEDYNNQMTWAQSCYVTGADGSESTRMLAAGS----- 243  
 DB 883 S-----VLTITTPSTINVR-----PIIEYVTTIDGNS--WIPRGSVITLT 922  
 QY 244 -----IFPPGNYDGLMDNDIALSGMAHKSYNVDGTGINSFTRIIIGKFSWYVINSGLDR 299  
 DB 923 QSVPIVEGKNEG-----SYNVSNGV-AITVNOPIITERF--VKQINSGFV 964  
 QY 300 GHAVIII 306  
 DB 965 GSVIII 971

RESULT 12  
 Q7V8S5\_PROMM PRELIMINARY; PRT; 1765 AA.  
 AC Q7V8S5;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Hemolysin-type calcium-binding region: RTX N-terminal domain.  
 GN OrderedLocustNames=PMW0256;  
 OS Prochlorococcus marinus (strain MIT 9313).  
 OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococcaceae;

OC Prochlorococcus.  
 NCBI\_TaxID=74547;  
 RN NUCLEOTIDE SEQUENCE.  
 RP MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;  
 RA Rocap G., Larimer F.W., Lamerdin J.E., Malfatti S., Chain P.,  
 RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,  
 RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regala W., Shah M.,  
 RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
 RA Webb S.A., Zinner B.R., Chisholm S.W.;  
 RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
 niche differentiation."  
 RL Nature 424:1042-1047(2003).  
 DR EMBL; BX572095; CAE20431.1; -; Genomic DNA.  
 DR GO; GO:0005576; C:extracellular region; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0005529; F:sugar binding; IEA.  
 DR GO; GO:0009405; P:pathogenesis; IEA.  
 DR InterPro; IPR001343; Hemlysn\_Ca\_bind.  
 DR InterPro; IPR001304; Lectin\_C.  
 DR InterPro; IPR003995; RtxA.  
 DR Pfam; PF00353; Hemolysincbind; 5.  
 DR Pfam; PF00059; Lectin\_C; 1.  
 DR PRINTS; PR00313; CABNDNGRPT.  
 DR PRINTS; PR01488; RTXTOXINA.  
 DR PROSITE; PS00330; HEMOLYSIN\_CALCITUM; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 1765 AA; 187603 MW; 817CE4F5007580CD CRC64;

Query Match 6.8%; Score 121; DB 2; Length 1765;  
 Best Local Similarity 20.1%; Pred. No. 13;  
 Matches 75; Conservative 60; Mismatches 146; Indels 92; Gaps 14;  
 QY 4 REDSMWTCVSTFNVILELGLDNCNRYAN-GLNMIGVILINPTDEGNFVIDDVTYN 62  
 DB 705 RHISYGATGADGNSITSGSDSVNLQINAGSALGLNDSILSTSGNDITIDITAIIG 764  
 QY 63 DNIKIVYIDGSDIDGSDGFYTGPN-----EYNTIPNSQSYSLKSENSQITQ--IKR 115  
 DB 765 ENS-----FYKSSSSRSRSPDSTGSNSDYSRIISRSSSYNDYTN 807  
 QY 116 YVSCNSTRRLTKSPSAKVTTTSGKVIT-----QNSINSRVYVNAIDATN 163  
 DB 808 YSNSYERSDASSSPFNDRSTNSRGINQSTYKNSYSGYNGWYNNSSYERSHNLNSY--YD 866  
 QY 164 FTDELRTTETREPNQSY---TSHKSTNSLVYHWTLIPRSLKQNMWR--EDYNNQWT 218  
 DB 867 YSNSSSSSSHNSNSYSRYNDTRYSNSSDYLNSSHQSQSYDNNRQSSYDYDHS 926  
 QY 219 WAGCYVKTGADG-----GSESTR--WLAGSIFPPGN-----YDGLWLDNDIAL-- 261  
 DB 927 RHISYGATGADGNSITSGSDSVNLQINAGSALGLNDSILSTSG--NDITIDVN 983  
 QY 262 -----SGMAHKSYNVDGTGINSFTRIIIGKFSWYVINSGLDRGHAIVITIDQYN 311  
 DB 984 AFGENSFYNKSSSRSSFPDSTGNSSDYSISYRSSS-----YSN 1024  
 QY 312 KYRIIFHAGYENS 324  
 DB 1025 DYTNSYSNDHS 1037

RESULT 13  
 O52984\_ECOLI PRELIMINARY; PRT; 814 AA.  
 AC O52984;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein precursor.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;  
 RN NUCLEOTIDE SEQUENCE  
 RP MEDLINE=99412288; PubMed=10482530;  
 RA Duchy T.G., Steadher L.H., Manning P.A.;  
 RT "55 Plus Biosynthesis Genes from Enterotoxigenic Escherichia coli  
 115:H40";  
 RL J. Bacteriol. 181:5847-5851(1999).  
 DR EMBL: AJ224079; CAJ1822.1; -: Genomic\_DNA.  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR GO: GO:0005215; F:transporter activity; IEA.  
 DR GO: GO:0006810; P:transport; IEA.  
 KW Hypothetical protein; Signal.  
 FT SIGNAL 1 21 Potential.  
 FT CHAIN 1 1 hypothetical protein.  
 FT CHAIN 22 814 hypothetical protein.  
 SQ SEQUENCE 814 AA; 90322 MW; C2108P2461BD892 CRC64;

Query Match 6.7%; Score 119; DB 2; Length 814;  
 Best Local Similarity 20.0%; Pred. No. 7.2;  
 Matches 94; Conservative 62; Mismatches 141; Indels 172; Gaps 22;

QY 1 FTLEDSMSDWTGVSTFNV-----ILFTGLDNCNIVA-----NG 34  
 DB 267 FGTAEBLVDRSYGKTFSSYSPSTGIVEGKDNILVYALTAQAGYEINLANLPYQYNA 326  
 QY 35 LNM-----GIYINITPTDE-----GPFVIDDVTLDNRIKVD----- 69  
 DB 327 LVQVSSSGIVSSQNLINNTGSPNSDFSMHLFVGNSSGSDNEFVKTEVEIESGVLP 386  
 QY 70 -----YIDGSDIDGSGMFYTGPNBEVNTI-----PNSQSYSLKSENSQITQ 112  
 DB 387 VNTLIALYVGAKVD-----KTIYISGLMFOKEPISVLRKGGGQGFHYE 433  
 QY 113 IKRVSCSNTSRLT-----KSFSAKVTTSKGVISITON-SINSSRVVINALDAT 162  
 DB 434 MKSELEBLSLMKKTSYSGKMNGLKSSDTMTLSAGYNFVMSVSNVGVYSSWRPD 493  
 QY 163 NF--TDELRTTETREPNOSYTHK-----SSTNSLYHTWTITPISLQNRME 211  
 DB 494 YFYANTDHLMESEFRRKKNYSNKL YANMYNFPGNSLYLNTY-----KELRGN 545  
 QY 212 DY-----NNGMTWAQSCYKTDGSGEST-----RWLAAGSIFPPGNY 250  
 DB 546 DYGSVSLGMITSLGNRSF---NSGFYKNGADITMSSTVDYAKSLDMWSVSV----- 596  
 QY 251 DGLWLDND---IALSGNAHKS-----YNDVTGINQLSF---TRIIGK-----GF 288  
 DB 597 -GRFSDSYNSATYSLSHNSNEVRGAGYYATDNGSQLTLTFADSTQIINSNGIYFTSS 655  
 QY 289 SWTYN--ISGLDRGHAVIIIDYGNKRYLLFHAGYENSDEPLSSIVY 334  
 DB 656 SWKNAPFIRGDKAKYDISVRNMTDNTTR-----YPSDNTIISVEVY 698

RESULT 14  
 O832P7\_ENTPA PRELIMINARY; PRT; 893 AA.  
 ID O832P7;  
 AC O832P7;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Conserved domain protein.  
 GN OrderedDomainNames=EF2174;  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=V583 / ATCC 700802;  
 RX MEDLINE=22550857; PubMed=12663927; DOI=10.1126/science.1080613;  
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,  
 RA Read T.D., Fouts D.B., Eisen J.A., Gill S.R., Heidelberg J.F.,

RA Tettelin H., Dodson R.J., Umayam L.A., Brinkac L.M., Beanan M.J.,  
 RA Dougherty S.C., Deboy R.T., Durkin S.A., Kolony J.F., Madupu R.,  
 RA Nelson W.C., Vamathevan J.T., Tran B., Upton J., Hansen T., Shetty J.,  
 RA Knouri H.M., Uetrich T.R., Radue D., Ketchum K.A., Dougherty B.A.,  
 RA Fraser C.M.;  
 RT "Role of mobile DNA in the evolution of vancomycin-resistant  
 RT Enterococcus faecalis";  
 RL Science 299:2071-2074(2002).  
 DR EMBL: AE016953; MA081906.1; -: Genomic\_DNA.  
 DR TIGR: EF2174;  
 DR GO: GO:0003796; F:lysozyme activity; IEA.  
 DR GO: GO:0016998; P:cell wall catabolism; IEA.  
 DR GO: GO:0009253; P:peptidoglycan catabolism; IEA.  
 DR InterPro: IPR02053; Glyco\_hydro\_25.  
 DR InterPro: IPR010916; TONB\_Box\_N.  
 DR Pfam: PF01183; Glyco\_hydro\_25; 1.  
 DR ProDom: PD004620; Glyco\_hydro\_25; 1.  
 DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 893 AA; 99579 MW; BCEFD9764AE80DB CRC64;

Query Match 6.7%; Score 119; DB 2; Length 893;  
 Best Local Similarity 19.4%; Pred. No. 8;  
 Matches 74; Conservative 58; Mismatches 119; Indels 130; Gaps 20;

QY 21 LFTGLDND-----CNVANGLMIGVIINITPTDEGNFVID-----DVTLDNRIKVD 69  
 DB 138 MKNGLSRVARATVAVYANDPVLPG-----KNITDVSMMNGDISVAEYQIKS 186  
 QY 70 Y-IDGSDIDGSGMFYTGPNBEVNTI-PNSQSYSLKSENSQITQIKRYVSCSNTSRLTK 128  
 DB 187 YGVTVGSVKLTCTETVY-NFYAAGQIRMAKAALYKS-----AHYSYVSAA-TAQDEAR 240  
 QY 129 SPSAKVTTSKGVISITONSINSSRVVINALDATNFTDELRTTETREPNOSYTHKSS 188  
 DB 241 YF-AQAANSNG-----LDKNTIMFNDAPPTLTNNRMAHNSVAFNODLKALGYK 290  
 QY 189 TNSLYHTWTI-----PSLSLQNRWEDYNNGMTWAQSCY 225  
 DB 291 NDALYVGKMWLTNGYIDTSAFGRDVAQYPTTPDS---SMQNDHGAMQWSQMTF 346  
 QY 226 KTGADGSGESTRWLAAGSIFPPG--NYDGLMDNDIALSGMAH-----KSY----- 269  
 DB 347 -----PGLANYEGRPPISTWYSFLAMGSSGSDLSKYTYTNP 385  
 QY 270 -----NVDGTINQLSF-----TRIIGFSWTVNISGLDRGHAVITIDQ---YG 310  
 DB 386 GRVIMKNDTFYQDVAFRTPGWRKKNLTVTIRGIR--YSSAGIPR---LVTDQGYLTA 439  
 QY 311 NKRYLLFHAGYENSDEPLSS 331  
 DB 440 NKDYVL--AAQSNIDLYFTTN. 458

RESULT 15  
 Q4L555\_STRAJ PRELIMINARY; PRT; 1361 AA.  
 ID Q4L555;  
 AC Q4L555;  
 DT 13-SEP-2005 (TREMBlrel. 31, Created)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)  
 DE Autolysin.  
 GN Name=act1; ORFNames=SH1911;  
 OS Staphylococcus haemolyticus (strain JSC1435).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=279808;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=JSC1435;  
 RA Takeuchi Y., Katanabe S., Baba T., Yuzawa H., Ito T., Cui L.,  
 RA Morimoto Y., Kuroda M., Takahashi M., Anai A., Baba S., Fukui S.,  
 RA Lee J.C., Hiratake K.;  
 RT "Whole genome sequencing of Staphylococcus haemolyticus uncovers



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 30, 2006, 09:40:03 ; Search time 128.372 Seconds  
(without alignments)  
5726.190 Million cell updates/sec

Title: US-09-889-874a-23

Perfect score: 8879  
Sequence: 1 VYIKFLKLFRRITMSDNNP.....PKIIIGREKTVKPTFRP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues  
Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*  
1: geneseqp1980s.\*  
2: geneseqp1980s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*  
9: geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8879	100.0	1673	3	AAV95707
2	1638	18.4	1564	2	AAV95707
3	1601	18.0	1410	8	ADQ26337
4	1583	17.8	1787	6	ABM67433
5	1569	17.7	1590	6	ABM67283
6	932.5	10.5	982	7	ADG73113
7	932.5	10.5	982	7	ADL12168
8	517.5	5.8	1317	8	ADS22210
9	432	4.9	2234	5	ABG31849
10	386	4.3	2234	6	ABU18641
11	365	4.1	1400	7	ADCO1365
12	361	4.1	1510	7	ADG30698
13	353.5	4.0	843	6	ABU48290
14	350	3.9	1515	6	ABU19676
15	342	3.9	1397	6	ABU15135
16	341.5	3.8	1404	7	ADCO0960
17	341	3.8	1394	7	ADCO1426
18	340.5	3.8	1411	6	ABU15136
19	339	3.8	1426	6	AAH5983
20	339	3.8	1426	6	ABU14693
21	331.5	3.7	1377	4	AAU34791
22	331.5	3.7	1377	4	AAAG8997
23	331.5	3.7	1377	6	ABU14807
24	328	3.7	1572	6	ABU41491

25	324.5	3.7	1429	6	ABM69806	Abm69806
26	322	3.6	1385	6	ABU40318	Abu40318
27	317.5	3.6	1504	6	ABM6973	Abm6973
28	317.5	3.6	1565	6	ABU16634	Abu16634
29	317	3.6	998	7	ABO79678	Abu79678
30	314	3.5	1395	6	ABU22662	Abu22662
31	313	3.5	1627	6	ADA35317	Ada35317
32	307.5	3.5	932	6	ABU16693	Abu16693
33	306.5	3.5	1530	6	ABU40251	Abu40251
34	306	3.4	1566	7	ADF05044	Adf05044
35	298.5	3.4	1531	6	ABU21384	Abu21384
36	298	3.4	885	4	AAU52772	Aau52772
37	298	3.4	885	6	ABM49291	Abm49291
38	298	3.4	920	6	ABM65127	Abm65127
39	297	3.3	1512	6	ABU50462	Abu50462
40	294	3.3	1439	7	ADF07275	Adf07275
41	292	3.3	1626	7	ABO75059	Abu75059
42	291.5	3.3	1253	6	ABU40990	Abu40990
43	291	3.3	2346	5	ADL16955	Adl16955
44	290.5	3.3	2802	5	ADL16953	Adl16953
45	289.5	3.3	1364	6	ABU47385	Abu47385

## ALIGNMENTS

RESULT 1  
ID AAV95707 standard; protein; 1673 AA.  
XX AAV95707;  
AC AAV95707;  
DT 25-OCT-2000 (first entry)  
XX Cosmid cHRIMS encoded protein P14-2F.  
DE Cosmid cHRIMS; nematocides; biological control agent;  
XX Cosmid cHRIMS; nematocides; biological control agent;  
KW transgenic plant; helminthiasis; P14-2F.  
XX Xenorhabdus bovienii.  
OS WO200042855-A1.  
PN 27-JUL-2000.  
PD 24-JAN-2000; 2000WO-GB000219.  
PE 22-JAN-1999; 99GB-00001499.  
PR (HORT-) HORTICULTURE RES INT.  
PA Morgan JAW, Jarrett P, Ellis D, Ousley MA;  
XX WPI; 2000-499157/44.  
XX N-PSDB; AAA50029.  
DR Novel composition used to control parasitic nematodes, especially in a  
PT plants such as maize, cotton, soya, and rice, comprises a bacterium which  
PS is a symbiont of an entomopathogenic nematode.  
XX Example 6; Page 42-43; 74pp; English.  
XX The present sequence is that of protein P14-2F encoded by an open reading  
XX frame identified in cosmid cHRIMS (see AAA50029). cHRIMS was obtained by  
XX ligating Xenorhabdus bovienii strain I73 (NCIMB 40986) Sau3A-digested DNA  
XX fragments into the BamHI site of the Stragene cosmid vector Supercoiled,  
XX packaging into Escherichia coli XL Blue 1, and screening for nematocidal  
XX activity against Caenorhabditis elegans. Analysis of the DNA indicated a  
XX number of open reading frames for which the corresponding protein  
XX sequences were determined (see AAV95685-Y95735). Nematodes can be  
XX controlled through the use of bacteria associated symbiotically with an  
XX entomopathogenic nematode. Such bacteria include Xenorhabdus and  
XX Photorhabdus spp. such as X. bovienii strain I73. The symbiont bacteria,

CC an engineered bacterium, or a nematocidal protein obtained from such  
 CC bacteria, particularly p13-1f (see AAY95706) or p14-2f can be used to  
 CC control helminthiasis in a human or domesticated animal or for the  
 CC control of plant pathogen nematodes. Also claimed are vectors for  
 CC expressing nematocidal proteins in host cells, and transgenic plants  
 CC

XX Sequence 1673 AA;

Query Match 100.0%; Score 8879; DB 3; Length 1673;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1673; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYIKFLKFRITMSDNNNEFTQANNFTSAVSGVDPRGLYNIQITLGHVGNGLGPT 60  
 DB 1 VYIKFLKFRITMSDNNNEFTQANNFTSAVSGVDPRGLYNIQITLGHVGNGLGPT 60  
 QY 61 LPLTLVSPLNKTDIGFGIGFNLGYDRKNSILSLSTGENYKVIETDKTVKLOQKLD 120  
 DB 61 LPLTLVSPLNKTDIGFGIGFNLGYDRKNSILSLSTGENYKVIETDKTVKLOQKLD 120  
 QY 121 NLREKOLKENCYRIHKSGLIEVLTGPNNAFDLAKPKKLLNPAHAITYIDMNFETOP 180  
 DB 121 NLREKOLKENCYRIHKSGLIEVLTGPNNAFDLAKPKKLLNPAHAITYIDMNFETOP 180  
 QY 181 RLNRIDYDLGDHDIPLNLBYOGLIKTILFLPGQKGYRTELRLNROLNSIHNFSLGN 240  
 DB 181 RLNRIDYDLGDHDIPLNLBYOGLIKTILFLPGQKGYRTELRLNROLNSIHNFSLGN 240  
 QY 241 ENPLTWSFGYTPIGKNGILGOWITSMAPGGLKETTYSNNOQHHPQSANLEFVLYYT 300  
 DB 241 ENPLTWSFGYTPIGKNGILGOWITSMAPGGLKETTYSNNOQHHPQSANLEFVLYYT 300  
 QY 301 LMKOVPGAGOPAIQAEVSYTSHNVYGGSGNGIMNNKLDNLGLMTEYVYGSTESRRYXDK 360  
 DB 301 LMKOVPGAGOPAIQAEVSYTSHNVYGGSGNGIMNNKLDNLGLMTEYVYGSTESRRYXDK 360  
 QY 361 EGHQOIVARIERTYNNYHLITSECKQONGYIQTTETAYYAIIGHNPDSPQSFOLPKTKTE 420  
 DB 361 EGHQOIVARIERTYNNYHLITSECKQONGYIQTTETAYYAIIGHNPDSPQSFOLPKTKTE 420  
 QY 421 TWRADNSYRSEIETETTFDESIGNPLTKYIKDKTKTKIISPSTHWEYYPAGEVNCPEEP 480  
 DB 421 TWRADNSYRSEIETETTFDESIGNPLTKYIKDKTKTKIISPSTHWEYYPAGEVNCPEEP 480  
 QY 481 YGFRFVVKIITQTPDSEFKDPEKFIQYRYSLIGSOSHVLKIEERYYSATOLLNSTLF 540  
 DB 481 YGFRFVVKIITQTPDSEFKDPEKFIQYRYSLIGSOSHVLKIEERYYSATOLLNSTLF 540  
 QY 541 QYNTDKSELGRLLKQTECTKGSENGKTVSVHKFTYTKODDTLQOSSHITTHDNFTIHRSQ 600  
 DB 541 QYNTDKSELGRLLKQTECTKGSENGKTVSVHKFTYTKODDTLQOSSHITTHDNFTIHRSQ 600  
 QY 601 VRSRYTGLFSDTDTKDITVQMSYDKGRLLTRLTNSGTPYANTLTYYEYLNLODDNRP 660  
 DB 601 VRSRYTGLFSDTDTKDITVQMSYDKGRLLTRLTNSGTPYANTLTYYEYLNLODDNRP 660  
 QY 661 PVRSTYGLFSDTDTKDITVQMSYDKGRLLTRLTNSGTPYANTLTYYEYLNLODDNRP 720  
 DB 661 PVRSTYGLFSDTDTKDITVQMSYDKGRLLTRLTNSGTPYANTLTYYEYLNLODDNRP 720  
 QY 721 LTNQOQTDPDVKHLSMSKSYDNMQOIANTHWSYVSEKTIYVDTTLTATKOLQSNNSNV 780  
 DB 721 LTNQOQTDPDVKHLSMSKSYDNMQOIANTHWSYVSEKTIYVDTTLTATKOLQSNNSNV 780  
 QY 781 QTGKEVTTYPSQOPIQITLFDAGHLOSCHTLTGDMDRVAKETDAIGQCTIYQYDNYN 840  
 DB 781 QTGKEVTTYPSQOPIQITLFDAGHLOSCHTLTGDMDRVAKETDAIGQCTIYQYDNYN 840  
 QY 841 RVIOITLDPGTIVNKKYAPFSTDTLITDRVNGISLGGQTFPGLSRLTQSOQGGVWMAVT 900  
 DB 841 RVIOITLDPGTIVNKKYAPFSTDTLITDRVNGISLGGQTFPGLSRLTQSOQGGVWMAVT 900  
 QY 901 YSAGNDQCPSTVITPDGFIHYQOPELDDAVLQVANSNETQOFSYNPVTGALLKVAEG 960

DB 901 YSAGNDQCPSTVITPDGFIHYQOPELDDAVLQVANSNETQOFSYNPVTGALLKVAEG 960  
 QY 961 QSLTPIYPSGRKMNINDMKMSYLAFTLRGLNGYTDLTGTIOKISRDPHGRVTOIKD 1020  
 DB 961 QSLTPIYPSGRKMNINDMKMSYLAFTLRGLNGYTDLTGTIOKISRDPHGRVTOIKD 1020  
 QY 1021 SSIKTTLANDDLNRHIGSOVTDLATGSHMLTTFVFPDLNREIGRKLCDSSGHTLIDQSW 1080  
 DB 1021 SSIKTTLANDDLNRHIGSOVTDLATGSHMLTTFVFPDLNREIGRKLCDSSGHTLIDQSW 1080  
 QY 1081 LKTQOLANRIVKXNGVLOQTEQSYSTRRLNRYKCDGACFPDKYGHSLVTQNFYDIY 1140  
 DB 1081 LKTQOLANRIVKXNGVLOQTEQSYSTRRLNRYKCDGACFPDKYGHSLVTQNFYDIY 1140  
 QY 1141 GNITACHTTPADGTEDHATFPANPDPCCOLTEVHTHPMPDNIRLKYDAGRVINITD 1200  
 DB 1141 GNITACHTTPADGTEDHATFPANPDPCCOLTEVHTHPMPDNIRLKYDAGRVINITD 1200  
 QY 1201 NHGNTENFTYDTLGRLONGGSGYGYDPLNRLVSOKTDLDCSLYYRETMVNEVRNGEM 1260  
 DB 1201 NHGNTENFTYDTLGRLONGGSGYGYDPLNRLVSOKTDLDCSLYYRETMVNEVRNGEM 1260  
 QY 1261 IRLIRGETTIAQORASKVLLTGTDSQOVSILTSBKQNSQEAYSAYGKHKSTANDASIL 1320  
 DB 1261 IRLIRGETTIAQORASKVLLTGTDSQOVSILTSBKQNSQEAYSAYGKHKSTANDASIL 1320  
 QY 1321 GYNGERADPVSGVTHLGNRYSYDPTLMRFPDPDLSPPFAGGINPYSYCLADPINKSDP 1380  
 DB 1321 GYNGERADPVSGVTHLGNRYSYDPTLMRFPDPDLSPPFAGGINPYSYCLADPINKSDP 1380  
 QY 1381 SGHLSQWAMTGIQWINGLGLLTATGMAIAAGGIAAIAASTTTALAFALSVTSIDT 1440  
 DB 1381 SGHLSQWAMTGIQWINGLGLLTATGMAIAAGGIAAIAASTTTALAFALSVTSIDT 1440  
 QY 1441 SIYSALIEDSPKASSILGWSMGMAAGLAEBAIKGTGLATHLGAFADGENALLKST 1500  
 DB 1441 SIYSALIEDSPKASSILGWSMGMAAGLAEBAIKGTGLATHLGAFADGENALLKST 1500  
 QY 1501 SESSRIKMGVTRSLDREIVANESQVTKDSRGYTDNFMKGBOALIVHGDKDFLYHTE 1560  
 DB 1501 SESSRIKMGVTRSLDREIVANESQVTKDSRGYTDNFMKGBOALIVHGDKDFLYHTE 1560  
 QY 1561 GNKNGKGPYTRHPEGLVDYLDKNNIVDLTGGDKFVHLISCYKSSGAADKMAKTYINR 1620  
 DB 1561 GNKNGKGPYTRHPEGLVDYLDKNNIVDLTGGDKFVHLISCYKSSGAADKMAKTYINR 1620  
 QY 1621 PVIAYSNKPITISQGLARIEBKDFPKSTYHSYDPKIIILGRTEKTVAPKTFRP 1673  
 DB 1621 PVIAYSNKPITISQGLARIEBKDFPKSTYHSYDPKIIILGRTEKTVAPKTFRP 1673

RESULT 2  
 AAY33727  
 ID AAY33727 standard; protein; 1584 AA.  
 XX  
 AC AAY33727;  
 XX  
 DT 09-NOV-1999 (first entry)  
 XX  
 KW Photobhabdus luminecens 176 kD insecticidal toxin.  
 XX  
 KW Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.  
 OS Photobhabdus luminecens.  
 XX  
 PN W09942589-A2.  
 XX  
 PD 26-AUG-1999.  
 XX  
 PF 18-FEB-1999; 99MO-EP001015.  
 XX  
 PR 20-FEB-1998; 98US-00027080.



[illegible]

QY	476	CPPEYGGTRPVKKIIOPIYSEFODDEKPIQRY----	SLISGQ-----	SHVTLKIEE	526
Db	443	CPADPELGSRFKSVTQKSGPDAAOSVANKKIHYHQEPFTTGAUYKVEYKVSERETIDN			502
QY	527	RHYSATOLNLTLPFYNTDKSELGILLKQTEBCTKGKNGKTVSVHKPFYTKODPTLQOSH			586
Db	503	KIARFFSVNBP-----TSKSH-GSLAKITSVANNQ-----	QTVTFKYEVSSEXTNIA		551
QY	587	SITTHDNFTIHRSOVRSRYTGRLPFSDDTDKOIVTOMSYDKLGRLLTRLTNPGTPYANTLT			646
Db	552	TVTFGDGAMHESKNTSITYTHRQJRKVDVNVNHYITDQSDLLGRITIGQIIDPETABEIKRN			611
QY	647	YDIELANIQDNNRPFPVITTTDVANGNQLRNEFDGAGRHSOCLKSDG-----	DGKE		698
Db	612	YVYOYGGDENDPWP-VWIEVDSQCVARKTKYTDGMRCSIEBDDDDGAMGTSGIYQGYT			670
QY	699	YTIHQOYDEBQGRHHTSYSDVLTNGRQOTP-----	DKVHLSMKSYDNMQOIANTHS		753
Db	671	RKVLARQIDVUJQSLKEISNDMLN--ISANPLVRLATPLVTTKYTKYDGMGNLSTEYS			728
QY	754	YGVSEKTIVDPEITLTATK-----	QLOSNNSNVQTKSEVTTYTPSQEPLOITLTPDEAGHL		807
Db	729	DGRLEIHDPIITRITITQGVKGLGNLINOQNF-----	EQPASIKAVVPDGI		776
QY	808	QSCHTLPDGDWRKPEIDAIQCTIYOYDYNRVIOITLPDGTIVNRKVPFSDTDLT			867
Db	777	YSTRYRYDGGRTVETTERDAGSHATQIDYDFDIRVKKTLTDDGTLLESAYVSFSEHELIS			836
QY	868	DIRVNGISLGGOTPDGSLRLTOSQDDGCVMAVYTSAGNDQCPSTVYTPDQGEINHQYPE			927
Db	837	ALANVNGTQGLAVYDGLGRVISTVYGVGKTEYLVGPOGDK--	PISGITSHNKQNDVLYY		895
QY	928	LDDAVLQVANSNEITQOFSYFNVTGALLTAAVAGOSIAPPI--	YUPSGLTKMEN--INDMKKMS		985
Db	896	LGSVMSKTTGTDOQNFYHSTKGTGLLSA--SEGVSGQNTYSIFPSGVLOQRESTRKMPIS			954
QY	986	---YIMTLRGLNGYTDLTGTIOKISRDTGRVTOIKDSSIKTTLVNDLNRHIGSOVTD			1043
Db	955	SGEYLVTMSGILOHKOFSFGHNHYSYDAQRLVXTEQDAUYATPEYDNVGRLLITTYKD			101
QY	1043	LATGHMLTTVPEPGLNREIGRKCDSSGHHLLDIQOSLTKOQLANRIVKLVNGVQREBO			1107
Db	1015	TTSLSQLTWKIEYAPFREIRKSLSDFSIOV--ITLSTYKNNQISOQLTSSIDGVMKER			1073
QY	1103	YSYDSRNLANQKCDGACPTDKYGHSLVTONFTYDIYGINITACHTTFADGEDBATKPF			1162
Db	1074	YOYDNNQRLSOGYCEGBESPIDHGRVLYNOQIYHYDQGNIKRLDNTYRDGKET--	VDYHF		1133
QY	1163	ANPTDPCQLEVHHHTHPMDPNIRLKYDKAGRVINIT--	DNHGNTENFTYDTLGRL--ON		1216
Db	1133	SG-ADPTQLIRI-----	TSDKQOIEHSYDANG--NLTRDEKQOT--	LIYDQNNRLVQYWD	1182
QY	1219	GGQSV---YGDPLNRLVSO--	KTDPLDCELYEETMLVNEVRNGEMRLRLTGETTIAQ		1277
Db	1183	RLGNLVCSYOYDALNKLTAQVLANGVNRQ--	HYASGKVTNLQDGEALTWYASSDQORGH		1241
QY	1274	QRA-----	SKVLTJTGDSQOQSVILTSDKONLSQEAYSAGKKSXTANDASILGYNERADP		1322
Db	1242	QSARKGQSVYQYQIDHNSITYIASQENELMALSTYTPGRFSLT---	SSLPGALNGAQVDP		1299
QY	1330	VSGVTHLNGVRSYDPTLMRPHRTDLSLSPFAGGGINPYCYLGADPINSDDPSGHLSWQAW			1388
Db	1299	VTGWVYFLGNGRVFNPVLMRPHSDPSMSPFGRGGINPYTYCOGDPINRIDLNGHLSAGGI			1356
QY	1390	TGIGMGJTAGLLTAT--	CGMAIA-----	ANGGIAAALASTSTYALAFALSVTSIDTYSIV	1444
Db	1359	LGIYVGAIGITVIGVISLGAAGAISAGLIAAGAGALGALASTSTALVTATVIGLADSDIGIA			1418
QY	1444	SGALEDASPKRASSILGVWVSMGMAAGLASEAKISGCTKLATLHGAFAEDGENVALKST--	SE		1502
Db	1419	SAALSSEKDPKTSIGILNMISAGIGVLSGISAI-----	TFY-----	SSLVKSARSQ	1466
QY	1503	SSRIKMGVTRSJDREIVNEBGGVYIKDHSR			1532



Db 1007 LGVSPFIQOTSYLLPMGLHAEVGAQASAMD-RMPSGRAR-BETHDIRSGKKKXHYR 1064  
 Qy 988 WTLRGLNGTYDLDLTGTTQKISRDT---H-GVTOIKDSSITTYLNDLNRHISQVTDL 1043  
 Db 1065 YSLTGNLEGADIDGAHARSYETAHAHVGLIADAAVVTALAYDGLQRLCSWTARD- 1123  
 Qy 1044 ATGHMLTTFYFEDGLNEIGRKLCDSSGHTLIDIOGWLKTOQLNRYKLVNGVLRTEQY 1103  
 Db 1124 GRGHMLTTFEFDSLGRBTKTLTAAESAEATTSQEMYPNGQLHQRKRSBGKPCDETF 1183  
 Qy 1104 SYDSRNLNOYKCGACBPTDKYSHSYVTONFTYDIYGNITACHTFADGEDHATPEFA 1163  
 Db 1184 VYDARNLKOYAAASGPGPKAYGNALRGCKFEFDAPFNIRKCTTVADGSGENVGEVLF 1243  
 Qy 1164 NPFDPCQTEVYHATHPD--MPDNIRLKYKAKAGVINITDNHGTENTFTYDTGLN---QN 1218  
 Db 1244 NPADPCQTYKVTNSALDKGYPPEALRYKQDQGRLEP--DEAG--RRLSYDALGRARVEG 1299  
 Qy 1219 GGGSV-VGYDPLNLVSGKTDLT--LDCELYYRETMVNE--VRNG-----EMIRLLRT 1266  
 Db 1300 GGGSASTGYDAHDLVCCQRETSQMDHRLYRANRLVNEWMTRSGQAPGADDDRVRLVYA 1359  
 Qy 1267 GETIIAQR---ASKVLNLTGDSQGSYILTSQRKNTSQEAYSAYGKHKT 1313  
 Db 1360 AGSCAAVNEGGDGSVAALMGTGDKGSIVSGAEGQAKHYATPYG-HQSS 1409

## RESULT 4

ABM67433  
 ID ABM67433 standard; protein; 1787 AA.

AC ABM67433;

DT 20-NOV-2003 (first entry)

DE Photorhabdus luminescens protein sequence #530.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KM detection; food; gene expression; plant; animal; microorganism; toxin;  
 KM antibiotic; biopesticide; virulence factor; disease model; plague;  
 XX whooping cough.

OS Photorhabdus luminescens.

XX KO200294867-A2.

XX 28-NOV-2002.

PF 07-FEB-2002; 2002KO-IB003040.

PR 07-FEB-2001; 2001FR-00001659.

PA (INSP ) INST PASTEUR.  
 (CNRS ) CNRS CENT NAT RECH SCI.

XX Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunet F, Danchin A,  
 PI Buchrieser C;

XX MPI; 2003-148459/14.

PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 useful e.g. as therapeutic antimicrobials and agricultural pesticides.

PS Claim 2; SEQ ID NO 530; 1205pp; French.

XX The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that

CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins

XX Sequence 1787 AA;

Query Match 17.8%; Score 1583; DB 6; Length 1787;  
 Best Local Similarity 28.2%; Pred. No. 4,1e-101; Indels 294; Gaps 65;  
 Matches 511; Conservative 280; Mismatches 727;

Qy 22 TOANNFTSAVSGVDPRTGTYNIQITLGVNGNLGPTLPILSYSPINKTKTDIGFSGIF 81  
 Db 30 SQATNFTGAMGAGVDPRTGTFPTQIPYTGAGNDLMGPDLATITLVNPLTYLNTGYGLGF 89  
 Qy 82 NFGLSVDRKNSLSTSGENYKVIETDKY-----KLOQKLDNLRPEKDLKENCYRI 135  
 Db 90 SDNFTRYDTQYVTLATGSIYHREKANEVVDQAVTFHRAKPAHKKVKE--KDAFW 147  
 Qy 136 IHKSGDIEVLTFGNNAFDLKVPEKLLNPAGHAIYIDMNFETAPQ-----RLNRIYDLDG 191  
 Db 148 LYKTSREKLTOL-DRANPAVAVSEIYAPSGHLCVKNASFVNHNNTYQMLRECDAME- 205  
 Qy 192 HDIPLNLKXQGLIKTILLPFGQKSGRYTELRFPLNQLMSIHNFSLGNENPLTWSEGYT 251  
 Db 206 ---TLKADLATEKIEFTVWPSPESGYTVLNMNTDLDQTIVASAS-----ELTWLHLEY 257  
 Qy 252 PIGKNGILGQWITSMRPAKGLKXTVNSNNQGHHP----- 288  
 Db 258 TEGAHKKI---LKVTPPSGLIERVYV--HETGHTLTPKCIQYPPAMPGIIRDPKS 312  
 Qy 289 -QSANLPLVPLPYTLMKQVPAGOPALQAEYSYSHNVVGGSGNGINV-----NKLNDLYGL 343  
 Db 313 STTATVHFPPYVQHNIAGSGSPDYIRYFEPENFLGQKNKMDPIPLPOQDANAYLA 372  
 Qy 344 MTEYNVGSSTRYKQKGGHDIVRIBRYNNYHLLTSECK-----QONGYIQTETAY 397  
 Db 373 NSEYKVTSTEVREYNEKR-----YCHIREYKPHILVSETETVEVTPSRQDKETIIKY 427  
 Qy 398 YAIIGHNFD--SQSPQQLPKTETWRSADN-----YSEITETTFDSGNPLTVYIK 450  
 Db 428 YADVGSFDDNKPQPLMPNTVETIWHNPSSASTORREKETTQWYNNQCNISMTL 487  
 Qy 451 DKKTOKIISPSTHMEYVPPAGEV---DNCPPRPYGFPRFKI--IOPPYD-SERKDDPEK 505  
 Db 488 DNTTKT-----TYAPDGBETTDTHCPAENGFRIKEJAVEAPSPILTKITILIR 540  
 Qy 506 FIOYRSLIGSOS-----HYTLKIEBHYS-----ATOL-----INSTLFQYNTD 545  
 Db 541 KTYNYKSYDTYSPPKNNQASVVKSMVLPLSEHYHSRDCADLBEVKIKNTYSFYENTQ 600  
 Qy 546 KSELGRLLKQTECKGNG-KTYSVNHKFTYTKQDDTLQOQSHSITTHDNF-----TIHRS 599  
 Db 601 NAFANGVEQRNSYLPFENGQTRSITYEDYSWNSNKNKAGASCIKTKTSGSKCGTPVSH 660  
 Qy 600 QVRSRYTGRLESPDPTDITVQMSYDKGLRLTRTNSGTPRYANTLYVDLNNLLODNR 659  
 Db 661 QVRSRSTGRLLPKQADNNNTVQYDTIGHLISSTINADTAYTEKTVKAMSYSNKR----- 716  
 Qy 660 PPVYITTDVNGQLRNEFDAGRHVSQCLKSDGDKFTYTHIQOYDEGRRHHTSYSD 719  
 Db 717 --VTVQTDIHERYITTEMGRLRPLKKSYPAGYQCKQDMERYQYNPLGQCOALISCD 774

```

QY 720 YLTNGROOTDDBKVHLSMSKSYDNMGOLANTHMSGVSEKLTVPITLTATKQLOSNNN 779
Db 775 HML-FENKSEKCKSMYVSLSEYDMGHNSRLRYDGDGTSVKATPPIKMTSHQVSNDES 833
QY 780 VQTGEVTTYTPSOQPIQITLFDAGHLQSCHTLLTRDGMWRKRETDALIGQCTTIYQYDNY 839
Db 834 QSSGIIRTTYNQFOGITATERLTLSRIQOGCMHYLRDELGR-L-VSINANGWTLTLLAYDAF 892
QY 840 NRVIQITLPDGTIVNRKKA-PPSTDTLTITLRV--GISLGOQTFDDGSLRSLTQSDGGRV 896
Db 893 DRVIKQTFADDTITSMAYENGVSVRMSATPLGINQTPILGTQILDGIGRATIMESGGRK 952
QY 897 WATYASGNDQCFSPVITPDCQ-----FHYQOPELDDAVYQVVS--NEITQOFSYN 947
Db 953 IKLDYE-GASFPVDTIVYVKGPTSGDKDVIHAYEPEPLNNAVYKIRAGDDVDVQTRKYD 1011
QY 948 PVTGALLKAVAE-----GQSLTPITYPSGRLLKME-----NINDMKMSYLMWLR-G 992
Db 1012 PKTGLTTEALIEYKXNDKEMISKLIIPDYTLSGOLASHKLVSYNNKNNRAYMHFPCYSFS 1071
QY 993 LENGTYDLTG-----IQKISRDTGRVTOIKDSSKITLANTDULNRHIGSOVTL-- 1043
Db 1072 PANPPTAKSTINGRXYIINKYSYDNQGRLETRNNNDIEIMLKYDKLSRHVCSQSTEXLFD 1131
QY 1044 ATGHMLTTTVEFPDGLNREIGRK-----LCDSSGHT-LDIOQSWLKTQOLANRLVYKLVNGV 1096
Db 1132 TINNKITTSILDLDPGRRETERTIRRHNGLRKNNKGTIDHISQTRPEQDKINKKL-LHGT 1190
QY 1097 LQ-RTEQYSYDSRNRLNQKCDGAECPYDKYGHSLVTO--NFYDIYINITACHTTFADG 1153
Db 1191 KQISEVYVYTHKRGLETTYTL-----MELVGEREITQCSYKDYKDLGNITQ--HSITLEG 1243
QY 1154 TEDHATPEFANPT--DPCQLTEVNH-THPDMPDNIRLKYDRAKGVINITD--NHGNTENFT 1209
Db 1244 KITSTYYVNGNITDPCQLIDVSTCTGNSSSRFTYFNORGALVCEINDNNKIRITWT 1303
QY 1210 YDTIGRLONGQGSV-----YGYDPLNRLV--SOKTDT-LDCELYARETMLVN----- 1253
Db 1304 YDSIGRIITTYDALHAKVETRIPLFDATNMLIKRSKSKGTPIPHHDLSTYSSNSLVHDNRYG 1363
QY 1254 EVRNGEMIR-----LLRTGETTIAQ--QRAKSVLLTGTDSQOQSVILTSDKONTLSQE 1302
Db 1364 EKRDMAADRKNNKVVGICLGFSPQTPCHQTPSVASRYETETATDQGSVIATFOGSDVQHI 1423
QY 1303 AYSAYGKRKSTANDASITIG-----YNGERADPVSGVTHLGNRYSTDPITLM 1348
Db 1424 AYSFPG--VTTQAMVTVAGQOPPHNTAIEBPRFNGEOMDTASASYLLGNGRRAVRPDLM 1480
QY 1349 RFHTPDSLSPFGAGGINPYCYCLADPIRSPDSGHLNQAWTGMGIAGLLLTATGGM 1408
Db 1481 RFTAPDSMSPFGAGGINAYVCGSDPVNLNDPSGHSIGMGWANNITTGIGILLAPFTYGG 1540
QY 1409 AIAAAGIAAIASTSTTALAFAALSVTSDITSIVSALGALBPASASSILGVMSMGMAA 1468
Db 1541 SLEIGLGVNAARGLT-----ALDAAISGVTAIASGALENKNPETSRLGLMWSIGLC-- 1590
QY 1469 GLAESAITGCTKLA-----THLGAFA-----AEDGENALLKSTSE--- 1502
Db 1591 --LPSMVVIGYSILAQMVVNVRLTNSFRTPHYFTSLGEVNLSSKSSDWNVARSINSGENW 1648
QY 1503 -----SSRIKVG--VTRSLDREIVRNEGOVIVDHSRGVYTDNMGKGEQAI-----L 1547
Db 1649 HSEVGLNGRTTWGSDTKIRGLD--IKYPLEQISRRPSNG--DIYLLSGSHGVONGDWWL 1703
QY 1548 VHGDQDGLYH-----TEGNKRNKGAPYTRHTPRLQVLDYLDKNNIIVDLTQGDGPV 1598
Db 1704 INGSRRGSLHPPFKSDMTVYGSGWKGRGTIVVRN-----LATMSEIDPGLTLNNGNHI 1758
QY 1599 HLLSCTYKSSGA 1610
Db 1759 ILGYCYGRNDA 1770

```

```

RESULT 5
ABM67283
ID ABM67283 standard; protein; 1590 AA.
XX
AC ABM67283;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photobacterium luminescens protein sequence #380.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photobacterium luminescens.
XX
PN W0200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002MO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunet F, Danchin A;
PI Buchrieser C;
XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photobacterium luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 380; 1205bp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photobacterium luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 1590 AA;

```

Query Match 17.7%; Score 1569; DB 6; Length 1590;  
Best Local Similarity 30.8%; Pred. No. 3.3e-100;  
Matches 477; Conservative 236; Mismatches 636; Indels 200; Gaps 57;

```

QY 63 LTLSSPLANKTIDIFGIGFNFGLSVYDRKNSLSTSGENYK---VIETDTVKLQCKKL 119
Db 13 LNLSPFLTTIANFGIGWFRSLTMDVKTITFRSNGEGRKCKPLPNNNDISFKDKKL 72
QY 120 DNLSFEKDLKENCYRIHKSQDIEVLTFGNNAADPLKVPKTLNPAHAITYIDNFEATQ 179
Db 73 KDLRVYK-LDSNTFYVYVNNKNGIIEFLKRIGSS--DIAKTVALFPPDG----- 116

```

QY 180 PRNRYDLDGHDIDPLNLEYGLIKITLT--FPGKGEYRELAFL-NROLNHNFS 237  
 Db 117 ----EYFDLYNRPALSEIKRMTGKTLYKLNYSGNN---CTSVETPDNNLSAKIAFD 169  
 QY 238 LGNEMLPWSFGYTPIGK-----NGILGOMITSMAPGGLKETEVYNNNOG 284  
 Db 170 YRNDYLITVVPDASPIRISARPKMTYOTLKGI-F-PVISAFTPTGYVLEVYKEN--G 226  
 QY 285 HNFPOSANLEPVLRYVTLMKQVPGAGPAIOAERYSTY-S-HNYVGGSGNGI--WNNKLDNLVY 342  
 Db 227 HKYTDIESIPYAALITQ---PENGOPALISKSEYSESVHNFLCYSGRTSPDSQNLVYL 283  
 QY 343 LMTEVNYGSTESRRYKDKEGHDQIVRETYNNYHLLTSECQONGIQTETRAYAALIG 402  
 Db 284 VYGYKYTYSIE---RVLNGQONVISTYERVPDKFLMTKEAKTQDNKRILITETYNEDPS 339  
 QY 403 HNFDSOPSPQFLPK-----TKETWRSADNSYRSLETETTPESGNPLTKVTKDKKTK 456  
 Db 340 KFSFSEBPENIQPSHYLTRITDQOT-----NTSRESEVNKSDDMGNTLL--TETSIGIQ 393  
 QY 457 IISPTHEWYYPAGEVDCPPBPYGFTRPVKKIIQTPYDEKDDPEKEIQYRYSLIGS 516  
 Db 394 -----EYVYYPVNGEGNNCPADPLGFSRFLKSVTQKSPDAQSVANRVTSYTGKLP 447  
 QY 517 QSHVTLKIEERHYSATQNLSTL---FOY---NTDSELRLLKQTECTGSENGKTYSV 570  
 Db 448 FTGAIVK--EYVAKASSETIDSKVTRFNYPNSPNNSH--GLAKITYSVMNQ-----QTV 499  
 QY 571 HKFTYTKODDTLQOSSHITTHDNFTIHSQVRSRYTGRLEPSDPTDQIYVOMSYDKLGR 630  
 Db 500 TTFKYSSEBEMTNTNVTVPDGTNHSKRVTSIYTHROLRKDVNHNVTIDQSYDLSGRI 559  
 QY 631 LRFLLNSGPYPANTLYDYVELNLODNRPVYITTTDVNGNOLRNEFDGASHVSC 690  
 Db 560 IGGIDPBTKEIKRSYIYQPGDENDPWP--VMIEIDSGIRKTHYDMGRICTSEBQ 618  
 QY 691 DSDG-----DCKFTIHTQYDEGRHHTSTYSYLTN-----GROOTDDKXHL 737  
 Db 619 DDGVMGTSGIYQGYRKYRKLARQYDVLGQVKEISNDMLDLSANPLTRLTTP--LVTTK 676  
 QY 738 SKSYDNMGQIANTHWSYGVSEKITVDPITLTATK-----QLOSNNSNNVOTGEVTTYTP 791  
 Db 677 TYQYDGMGRKYSTEBISGRILEIHDPITRTITQYVKGGLMAYIQNNP----- 725  
 QY 792 SQOPIOTTFDEAGHLQSCHTLFRDGMWRVREKETAIGQCTIYOYDYNVNIOTLPDGT 851  
 Db 726 -EQPASIKVYVPGAIYSTRYVDGGRVTEITDAGVATQIEYDLFRIYVKTLPLDRT 784  
 QY 852 IYVRKTAPESTDILDIRVNGISLGQOTDGLSLTQSDGGRWAAVYTTSAQDQCPST 911  
 Db 785 ILSSAAYASFHEELISALNNGTQGLSLVYDGRVTRDVGRKTEYLLGSGQDK--PIQ 843  
 QY 912 VITP--DGGVIHYQYQPELDAVLQVANSNEITQOFSNPTVGLIKAVAGOSLTP1--YY 968  
 Db 844 SVTPAHKKOIIDLY--ALGSVMSKTTETTSQONFSTYQTKGALLSH--TEGVOSQNSY 900  
 QY 969 PSGRLEKMENTIN--DMKMS---YLMTLRGLNGYTLDTGTIOKISRDPHGRVTOIKDSIK 1024  
 Db 901 PEGVLGHESFSDNKRPISSGDYRTWMSGLIORKHDSFAHDHYVSYAEGSLVTEBSSQY 960  
 QY 1025 TITANVDLNRHISQVTLDTLGMHTTTFVFDGLNREIGKLCDSGHTLIDIOOSMLKIQ 1084  
 Db 961 ATFEYVNGVGLITTTTKDTTSLQLATKIEYDVFDRKIKSLSDPSIQV--ITLSYTKN 1019  
 QY 1085 QLANRIVKNGVLORTQOYVSDSRNRLNOYKCDGAECPCTKYGHSITQOFTYDIQNT 1144  
 Db 1020 QISORTISIDGVKMKERYQYDSNQRISOYQCEBEGSPVDHTGRVLSQLYHNDQMNK 1079  
 QY 1145 ACHTTFADGEDHATKFPANPTPCQITVHNHTHPDMNIRLKYDKAGRVINIT--DNHG 1203  
 Db 1080 RLNDYTRDGKET--VDYHFSQ--ADPTQIIRI-----TSDKQIEISYDANG--NLTDEKG 1130

QY 1204 NTEFTYDHLGRL---ONGGSY---YGYDPLNRVLSO--KTDTLDCLEYYRETMVNEY 1255  
 Db 1131 QT--LIYDQNNRLVQYKDSKNUVCQYQYDALKKLTAQVLANKTVRQ--YIASGVANYQ 1187  
 QY 1256 RNGEMIRLL-----RTGETIIAQORASKVLLTGTDSQOSVILTSRKONLSQEA 1303  
 Db 1188 LGDETITWLSDKQRLGHQSTKNKESVYYQ-----GTDHNSYVIASQNEIMLAIS 1239  
 QY 1304 YSAVGHKSTANDASILTYNGERADVPYGVTHLGNQYRSDPTLMFPHRPDISPPGAGG 1363  
 Db 1240 YTPYGRSLI---SSLPJGNGAQVDPVTWYFIGNCYRVFNPLMFPSPDSWSPGRGG 1296  
 QY 1364 INPYSYCLDPINRSPSGHLSQWAMTGIQMGJAGLLTAT--GGMAIA---AAGTIA 1417  
 Db 1297 VNPFTYQGGPINKRIDNGLHLSAGTILGIVLHIGIIVGVSYGAGALISAGIILAGGAL 1356  
 QY 1418 AAIASSTTALAAGALSVTSIDTISYVSGALBEDAPASSILGVNMGMAAGLAESAI-- 1475  
 Db 1357 GAIASSTAPAVRATVIGLAADSIGIASALSEKDPITAGILNISTGVLSPGISAITF 1416  
 QY 1476 -----KGGTKLANH--LQAF-AEDGENALIKSTSESRITKWGVTRS 1513  
 Db 1417 TSSLIKSARSQSVASTSVIGSVPIEFGEIA-----SRSSR--RMDIALS 1459  
  
 RESULT 6  
 ADG73113  
 ID ADG73113 standard; protein; 982 AA.  
 XX  
 AC ADG73113;  
 XX  
 DT 11-MAR-2004 (first entry)  
 XX  
 DE Pseudomonas syringae pv. tomato DC3000 Avr/Hop protein #53.  
 XX  
 KM Avr; Hop; transgenic plant; disease resistance; cancer; bacteria;  
 XX metabolic pathway; eukaryotic cell death; programmed cell death;  
 XX cytosstatic.  
 OS Pseudomonas syringae; pv. tomato str. DC3000.  
 XX  
 PN US2003204868-AI.  
 XX  
 PD 30-OCT-2003.  
 XX  
 PF 12-FEB-2003; 2003US-00365742.  
 XX  
 PR 12-FEB-2002; 2002US-0356408P.  
 XX 10-MAY-2002; 2002US-0380185P.  
 XX  
 PA (COLL/) COLIMER A.  
 PA (ALFA/) ALFANO J R.  
 PA (CARF/) CARTINHOUR S W.  
 PA (SCHN/) SCHNEIDER D J.  
 PA (TANG/) TANG X.  
 XX  
 PI Collier A, Alfano JR, Cartinhour SW, Schneider DJ, Tang X;  
 XX  
 DR WPI; 2003-875735/81.  
 XX  
 DR N-PSDB; ADG73112.  
 XX  
 PT New nucleic acid, useful in imparting disease resistance to a plant or in  
 XX preparing a composition for treating cancer.  
 PS  
 PS Claim 15; SEQ ID NO 106; 209bp; English.  
 CC  
 CC The present invention relates to the isolation of Pseudomonas syringae  
 CC pv. tomato DC3000 Avr/Hop proteins, and the polynucleotide sequences  
 CC encoding them. Also disclosed are expression vectors, host cells, and  
 CC transgenic plants comprising polynucleotide sequences of the invention.  
 CC The polynucleotide and polypeptide sequences are useful in imparting  
 CC disease resistance to a plant or in preparing a composition for treating  
 CC cancer. The sequences may also be used to make a plant hypersusceptible

to colonisation by nonpathogenic bacteria, modify a metabolic pathway in a cell, cause eukaryotic cell death, and inhibit programmed cell death.  
 CC The present sequence represents a *Pseudomonas syringae* pv. tomato DC3000  
 CC Avr/Hop protein of the invention.

XX Sequence 982 AA;

Query Match 10.5%; Score 932.5; DB 7; Length 982;  
 Best Local Similarity 29.7%; Pred. No. 8.3e-56;  
 Matches 285; Conservative 142; Mismatches 357; Indels 175; Gaps 33;

```

QY 642 ANTLYTYEELNNLODDNRPPFVITTTDVNGNQLNEFDGAGRHSQCLKDSGDGKEYTI 701
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 31 SSTLGXYND-----DMNQ--CCTTTDDN-VQTYEYSDPIGSDVHK-----GPIQKT 74
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 702 HTQGYDEQGRHHSTYSYDLTNGRQCTDPDKVHLSMSKSYDNWQIANTHWSYVSEKIT 761
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 75 WKSGDEGR-----ISGRSET-----WNL-----FGKPRIR 103
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 762 VDPITLTATQLOSNNNVQTKGKVTYTPSQPIQITLPDEAGHLQSCHTLTBDGMDRV 821
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 104 ----TLTAGKTGRKRTISMRSRLTT---TEDEL-----SRQTFLYDGLGRC 144
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 822 RKETDAIGCCTIYOYDNNRVITQLTLPDGTIVNRKVAFFSTDLITDIRV---NGIS--- 875
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 145 TEGRDALQOSTLPSYDWMRSWVSTLADGSVYNSVAPQSSSELATMLEVHNGTTRTV 204
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 876 LGQOTFPGLSRLTOSQDGRWATYTSANGDQCSYITPBGQFIHQYQPELDVAVLQV 935
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 205 AGTKFPGLEKRVYQTKGDRVEQFNVDAGEMQ--PRSRTAGLDNINFTYTRALTDQIFSS 263
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 936 ASNETITQOFSYNPVTGALLKVAEGQSLTPYPSGRKMKENINDMKMSYLM-----T 989
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 264 TAPDETAKFYDXTKSARLIEATNQGTTRYVDVHNLQGTWNL--LGAMETRRQSS 321
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 990 LRGLNGYTDL----TGTQKISRDTHGRTVQIKDSIKTLLYNDLNRIHISQVTDLA 1044
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 322 LLGSRPKRTDLKKGAGAEATRYDYDLGRIRFINSQMLRTIIDYDLVGLQCKVATEDLQ 381
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1045 TGHMLTTFVEFDGLNREIGRLCDSGSHLIDLIQSMKLTQOLANRYKLANVLORTQYS 1104
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 382 AGTVIITDMEYDDGQETLRTQTASNOALTLTQTAVDGLKTRDLOQASSPLIHETFS 441
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1105 YDSNRRLNQYCKDGAECPTDKKXSIYQNTFYDIGNITACHTTTFADGETDHAETFEAN 1164
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 442 YDPRGRLLTVNYLSSLPREDLOREMTRQIFSPBELDNITTCQRTFDGTSEBAFAFKGS 501
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1165 P-----TDPQQLTEVHHTHPMPDNIRLKYPKAGRVINITNHNNTENFTYDLGRL--- 1216
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 502 PGDDKHKRCQLLSIAVTPPKRTPDPTFSYDANGN--QLKDBHGN--SLHYDSQSRLLQV 557
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1217 -QNGOG--SVYGYPRKLRLVSOQKTDLDCEL--YTRFMLVNEVRANGMILLRTGETII 1271
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 558 AETGAPISQRYRYGHNLVAIR--DGNSEILRFEGHQLSSTVOEDRTQYIHLGEPPL 616
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1272 AQQ--RASVYLTLCTDSQSVILTSDKONLSOAYSAVYGHKXSTANDASILIYNGERAD 1328
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 617 GQQIVDDAEQTLILLITDANQSVMEGFQOGLRKAVYSYGERHSEALLSTANGREAVRE 676
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1329 FVSGVTHLNGYRSYDPTLMRPHTPDLSLSPFAGGINSYCLADPPIRSDPSGH----- 1383
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 677 AANGVYLLNGYRAVNPILMRPHSPDPLSPFAEGGVNPFYCLGNPILALRDPFGHDASGQ 736
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1384 -----LSNQ-----AWTGIGMGA-----GLLITATNGMAIAAAGI----- 1416
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 737 TGRLRPDEGALPMQGGGDMGWVGIGVFTVLGVAATITALTGTTPVTPGTVLGI 796
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1417 -----AAAIASSTTALAFG---ALSTSDITSIVGALGEDASPKASILGWVS----- 1462
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 797 SMTASAAAIVSTVSGALIVGTALTAASTTANTVAIVNN-----DQTGREGGMLGIAAV 851
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 1463 ---MGMAAGIASESAIKGTGLA-----THLGAFAEDGENALLKSTSESSR 1505
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

DB 852 PVGLVFGAGAVVAVAAAAKVAANAAGTIGVRSYRIG-LAAGARRTISASASAR 909  
 RESULT 7  
 ADL12168  
 ID ADL12168 standard; protein; 982 AA.  
 XX  
 AC ADL12168;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE *Pseudomonas syringae* anti-cancer protein #53.  
 XX  
 KM cytostatic; gene therapy; Avr; Hop; cancer.  
 XX  
 OS *Pseudomonas syringae*, pv tomato DC3000.  
 XX  
 PN W02003068930-A2.  
 XX  
 PD 21-AUG-2003.  
 XX  
 PF 12-FEB-2003; 2003WO-US004450.  
 XX  
 PR 12-FEB-2002; 2002US-0356408P.  
 PR 10-MAY-2002; 2002US-0380185P.  
 XX  
 PA (CORR ) CORNELL RES FOUND INC.  
 PA (USDA ) US SEC OF AGRIC.  
 PA (UTNE-) UNIV NEBRASKA.  
 PA (UNITV ) UNIV KANSAS STATE RES FOUND.  
 XX  
 PI Collmer A, Alfano JR, Cartinhour SW, Schneider DV, Tang X;  
 XX  
 DR WPI: 2003-679632/64.  
 DR N-PSDB; ADL12167.  
 XX  
 PT New nucleic acid molecule, useful for preparing a composition for  
 PT treating cancer.  
 XX  
 PS Claim 15; SEQ ID NO 106; 284bp; English.

XX  
 CC The invention relates to novel *Pseudomonas* Avr and Hop genes, a sequence  
 CC that hybridizes with these sequences under stringency conditions  
 CC comprising a hybridization medium that includes 0.9 x saline sodium  
 CC citrate (SSC) buffer at a temperature of 42 deg C. The nucleic acid  
 CC molecule is useful for preparing a composition for treating cancer. This  
 CC sequence corresponds to one of the proteins of the invention.

XX Sequence 982 AA;

Query Match 10.5%; Score 932.5; DB 7; Length 982;  
 Best Local Similarity 29.7%; Pred. No. 8.3e-56;  
 Matches 285; Conservative 142; Mismatches 357; Indels 175; Gaps 33;

```

QY 642 ANTLYTYEELNNLODDNRPPFVITTTDVNGNQLNEFDGAGRHSQCLKDSGDGKEYTI 701
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 31 SSTLGXYND-----DMNQ--CCTTTDDN-VQTYEYSDPIGSDVHK-----GPIQKT 74
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 702 HTQGYDEQGRHHSTYSYDLTNGRQCTDPDKVHLSMSKSYDNWQIANTHWSYVSEKIT 761
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 75 WKSGDEGR-----ISGRSET-----WNL-----FGKPRIR 103
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 762 VDPITLTATQLOSNNNVQTKGKVTYTPSQPIQITLPDEAGHLQSCHTLTBDGMDRV 821
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 104 ----TLTAGKTGRKRTISMRSRLTT---TEDEL-----SRQTFLYDGLGRC 144
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 822 RKETDAIGCCTIYOYDNNRVITQLTLPDGTIVNRKVAFFSTDLITDIRV---NGIS--- 875
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 145 TEGRDALQOSTLPSYDWMRSWVSTLADGSVYNSVAPQSSSELATMLEVHNGTTRTV 204
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 876 LGQOTFPGLSRLTOSQDGRWATYTSANGDQCSYITPBGQFIHQYQPELDVAVLQV 935
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 205 AGTKFPGLEKRVYQTKGDRVEQFNVDAGEMQ--PRSRTAGLDNINFTYTRALTDQIFSS 263
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```



QY 936 ASNEITQFSEYNPTVGTALKAVALGOSLPIYYPSGRLKMENTINDMKMSYLM-----T 989  
 DB 264 TADERTAKFYDKTSNARLIEATNPQGRTRTRYDVHNOJTGSETDNL--LQANETHRQSS 321  
 QY 990 LRGLNGYVDL-----TGTIOKISRDPHGRVTOIKOSIKITLTYNDLNRHIGSVYDLA 1044  
 DB 322 LLAGRPKRTDILKKGGAAGAEFRVYDILGRIRFINGNLTATTDYDVLGQLCKVATBDLQ 381  
 QY 1045 TGHMLTTTVEFDGLNREIGRKLCDSSGHTLDIQOSMLKTQOLANRIYKANGVLOREYQS 1104  
 DB 382 AGGVIIIDMEYDQOGGELIRTOGASNOALTLTQTMVNDGLKTRDIQOAGSPULHETFS 441  
 QY 1105 YDSRNRNLYQKCDGAECPDKGHSIYTONFTYDIGNITACHTFADGNEHATFEKAN 1164  
 DB 442 YDRGRGLTVNLYGSSLPDELOREMTROTFSPELDNITLCOTRFDTGSERLAFTKGS 501  
 QY 1165 P-----TDPCLTEVHTHPDMPDNIRLKYDKAGRVINITDNHNTENTFYDTLGRU--- 1216  
 DB 502 PGDDKHKDRQLSLIATTPRKTPDPFYSVDANGN--QLNDEHGN--SLHYDSQSRLLQV 557  
 QY 1217 -QNGQG--SYGCDPLNRLYSQKTDITDCEI--YRETMALVNEVRNGEMIRLRTGETTI 1271  
 DB 558 AETGGAPIISQYRYDGNHQLVATR-DGNESEILREFEHQHLSSTVQEDQRTQYLHGEQPL 616  
 QY 1272 AQC---RASKVLLTGTDSQSVILITSDKONLSQASVAYGKHKSTANDASILGYNBERAD 1328  
 DB 617 GQGIUDDAEOTLLILITDANOSVMEFGQQLRKAIVSAYGRSEALSTAGNGEVR 676  
 QY 1329 PVSQVTHLNGVRSYDPTLRLRFTPLRSPGAGGINSYCYLADPIRSDPSGH----- 1383  
 DB 677 AANGWILGNGRAYNPLMRHSPDLSPFAEGVAPYCYGNPILALDPTGHDSGQ 736  
 QY 1384 -----LSNQ-----AMTIGMGIA---GLLITATGMAIAAGI----- 1416  
 DB 737 TGRLRPEDEGALPMQGGSDIMGWVGIGVETVLVAATITATLGTATVTGTGYLGI 796  
 QY 1417 -----AAAIASSTTALAFG---ALSVTSDITSYSGALBEDASPASSIIGVMS----- 1462  
 DB 797 SMTASAAAVSTVSTGALIVGTALTAASTANTVAIVNN-----DQTAGEVGGLGIAAV 851  
 QY 1463 ----MGWGAAGLAESAIKGGTKLA-----THLGAFAEDGENALLKSTSESSR 1505  
 DB 852 PVLGVGFGAGAVVABAVAAAKVAANAAGTIGVRSVGRIG-LAAAGRRITSSAASGAR 909

RESULT 8  
 ADS22210 standard; protein; 1317 AA.  
 XX  
 AC ADS22210;  
 DT 02-DEC-2004 (first entry)  
 XX  
 DE Bacterial polypeptide #11243.  
 XX  
 KM Recombinant DNA construct; transformed plant; improved plant property;  
 KM cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;  
 KM pathogen tolerance; pest tolerance; plant disease resistance;  
 KM cell cycle pathway modification; plant growth regulator;  
 KM homologous recombination; seed oil yield; protein yield; carbohydrate;  
 KM nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
 KM bacterial polypeptide.  
 XX  
 OS Bacteria.  
 PN US2003233675-A1.  
 XX  
 PD 18-DEC-2003.  
 XX  
 PF 20-FEB-2003; 2003US-00369493.  
 XX  
 PR 21-FEB-2002; 2002US-0360039P.

XX  
 PA (CAOY) CAO Y.  
 PA (HINK) HINKLE G J.  
 PA (SLAT) SLATER S C.  
 PA (CHEN) CHEN X.  
 PA (GOLD) GOLDMAN B S.  
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
 DR WPI, 2004-061375/06.  
 XX  
 PT New recombinant DNA construct comprising a promoter positioned to provide  
 PT for expression of a polynucleotide encoding a polypeptide from a  
 PT microbial source, useful for producing plants with improved properties.  
 PS Claim 1, SEQ ID NO 11243; 122pp; English.  
 XX  
 CC The invention relates to a recombinant DNA construct comprising a  
 CC promoter functional in a plant cell, where the promoter is positioned to  
 CC provide for expression of a polynucleotide encoding a polypeptide from a  
 CC microbial source. The invention also relates to a transformed plant  
 CC comprising the recombinant DNA construct and a method of producing a  
 CC transformed plant having an improved property. The plant is a crop plant  
 CC such as maize or soybean. The method of producing a transformed plant  
 CC having an improved property comprises transforming a plant with the  
 CC recombinant DNA construct and growing the transformed plant, where the  
 CC polynucleotide or polypeptide is useful for improving plant properties.  
 CC The recombinant DNA construct is useful for producing plants with  
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
 CC increased resistance to plant disease, better growth rate by modification  
 CC of the cell cycle pathway with plant growth regulators, increased rate of  
 CC homologous recombination, modified seed oil or protein yield and/or  
 CC content, improved yield by modification of carbohydrate, nitrogen or  
 CC phosphorus use and/or uptake, by modification of photosynthesis or by  
 CC providing improved plant growth and development under at least one stress  
 CC condition, improved lignin production or improved galactomannan  
 CC production. This sequence represents a bacterial polypeptide used in the  
 CC scope of the invention. Note: The sequence data for this patent did not  
 CC form part of the printed specification but was obtained in electronic  
 CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
 XX  
 SQ Sequence 1317 AA;  
 QY Query Match 5.8%; Score 517.5; DB 8; Length 1317;  
 QY Best Local Similarity 21.2%; Pred. No. 2e-26;  
 QY Matches 334; Conservative 194; Mismatches 505; Indels 541; Gaps 72;  
 DB 39 TGLYNIQITLGHIVGNGNLGPTLPLTL--SYSPLNKTDIGFGIG--FNGSLSYDRKNSL 94  
 DB 56 TGSYFYQYDLSIPKRG-----LPLTVSRSYNMDNRSGLFGSGWTFNFMNMLTVDNNGN 110  
 QY 95 LSLSTGENYK---VIEFDKTVKLQOKKLDNLRREKDLKENCYRIHKSQGEIVLTGFNNN 151  
 DB 111 VTVLGGDGHDTDTIAMPDGTYSRPLSVFDDL-----INKSDTYTLTKKDDQT 157  
 QY 152 AFDLKVPKKLLN--PAGHAIYIDMNPATQPLNRIYDDLGDHDIPLNLAYQGLIKTI 208  
 DB 158 KYNFSESGKLVNIVDKNGNQI---NFTYTGSEQLTVY-TPASGREL-ILAYDHNG----- 206  
 QY 209 LTLFPQKGYRRELRLFLNRQLNSHINFSLGNNPLTWMSGYRPIG-----KNGIIGQ-- 261  
 DB 207 -----HIISTDMGRWVSYSDYDQGLLQCKNPIGKLS 241  
 QY 262 -----WTSMTAPG---LKETVYVSNNOGHFPOSANLPLVPLVYTLMKQVPGAGQP 311  
 DB 242 YTYDENHMTSITDPGNNPMKNTY---DEKHHVLSQS----- 276  
 QY 312 AIOAERYSYSHNVYVGGSGNINNNKLDNLGLMTBYNYGSTRRYKXGEGHQIVRIER 371  
 DB 277 -----NSILNATY---TENYDS-ENRK----- 293  
 QY 372 TYNNVHLLTSECKQNGYIQTETAYAYAIIGHNFDGQPSQFOLPKTKTETWRSADNSYRS 431

Db 294 -----TTFD-----PFGNKTYSPBHW-- 313  
Qy 432 ELET-----TFDESNPLTKVIKDKTKOIIISPTHWEYYPAGEVDCPPRYG 482  
Db 314 ELNETNOGLYTIISAYDENGRIISVTNENSKTTLA-----YDANGNIK-TTNPILG 364  
Qy 483 FTRPVKIIQTPYSEFKDDEKFIQRYVLSIGSHVTLKIEBRHSAQOLNS----- 537  
Db 365 YSK-----SMTYDS-----KNLISQDELHRKTSPEYDNSNLKISDALG 406  
Qy 538 --TFOYNTDKSELGRLLKOTECTK-----GENGR-----TYSVHAKT 574  
Db 407 HEYFSD-----KGOYIGETDSNKKTATPSYNNNGOITTTDANGTSAFTYTVGRVT 462  
Qy 575 -----YTKQDPTLQOSHSTTHDNFTIHSQVSRYSYGRU-----FSDPTKDI 618  
Db 463 TKTDAGKNRYTFQYDALDNLISITD-----PMQOTTSTNTMLLEIKLVLOMLKVXQSI 516  
Qy 619 VTQMSYDKLGRLLRTIANSCT---PYA-----NTLVYDELANLQDDNRPPPVIT 665  
Db 517 LTTLIXLXKXRMQWVEXLGTNMPPLAIWEPROTOKGHKTSYDYPDLNRQ-----VS 568  
Qy 666 TTDVNGNQLRNEPFDAGRHVS-----QCLKDSGDKFYTITHTQOYDE 708  
Db 569 VTNALGKTRKRYALGNKISITNAYGKSTYSNSLNQVKTVMAMGK---VVRNRYDA 625  
Qy 709 QGRHHTSYSD-----YLTNGRQOQDPPKVLHSMKSYDNWQOLANTHWSYSEKIT 761  
Db 626 VGNLITTDENGHKINCYDLSNRQVSVTDLARKTRNKRYDAVGNKISITNAYGKSTRYS 685  
Qy 762 VDPITLVRATQLOSNMNVQGEVTTYTPSQPIQITLPEAGHLDSCHTLTDMGMDRV 821  
Db 686 Y-----NSLN-----QLVKVTDMGQVRYNYDAVGNLS----- 715  
Qy 822 RKETDAIGQCTIYOYDNYNRVIOITLPGDTIVNRKYPFSTDTLITIRVNGISLQOTF 881  
Db 716 --TTDANGRKTNVGYDSLNRQVSTITNALGKTRKRYDAVGNKISSTOANMR---LTKYSY 770  
Qy 882 DGLSRLTOSQDG-GRWATYTYSA-GNDCCPSTVITPD--GCFIHQY-----QELDAV 932  
Db 771 DLSRLVYKVTAMGQVARYTYDAVGN---LISTTDAKGKHTDYGDSLROVSITDPL 825  
Qy 933 LOVASNETTOQFSYNPVYGALLKAVBEGOSLPIYPSGRUKMENINDMKGSYLTMLRG 992  
Db 826 GRITARNK-----YDAVGNKISSTDEGKITSYGVDLNRLTKYSYDQKVSY----- 873  
Qy 993 LENGVTDLGTLOKISRDPHGRVYTOIKDSIKTTLNVDLNRH-----GSOV--TDLA 1044  
Db 874 -----NYDAVGNRLTMKOSHGTAYKDKLNRLLSVLNPQKVSITYNK 918  
Qy 1045 TGHML-----TTTVEFDGLNREIGRLCDSCHTLD---IQOSWTKTOOLANRYKL 1093  
Db 919 VGNHVKMTYPDGKITSYSDAVNRLIG--VIDSDHITSYSYAKNGKNTKMTNPGV--- 973  
Qy 1094 NGVLORTQOYSDSNRNLNOYKCCDGAECPTKYGHSIYONPXYDIGNYTACHTTTFADG 1153  
Db 974 -----KTB-ISTYDKANRL-----VELINKTTOVSSYKTTLDAGNRLKVDQBLAEG 1020  
Qy 1154 TED-----HATEKFNPTDPCQLTEVHHTHPMPDN--IRLKADAGRVINIT- 1199  
Db 1021 VESGDSSELKESQLLTTTYGY--DKLYRLTKV-----DYPNKTIVSYKXDSGNRISMTT 1072  
Qy 1200 --DNHGTENFTYDTLGRLONGQGSVGYDPLNRLVSQKTDTLDELY-YETFMVLNVEVR 1256  
Db 1073 NVDGISTISYKYDADQLOSGNISYSYDKKGNLIKRVANSTQPMYSYBE---ANLKL 1129  
Qy 1257 N-----GEMIRLART---GFTIIAQORASVULLTGDSQOOSVLT 1293  
Db 1130 NVSEFVSNTNTPKSSYNFEYDGNRIITKTITNGE---NAOSTKYVDLINSALPQVLTE 1185  
Qy 1294 SDKONTL-----SOEA-----YSAYGKHSSTANDASIL----- 1320

Db 1186 SDTKNTTCYTYGTHDLSMTNSENABEYHYHVDGLSVRSLSDSKGIKNTLYLDYAFQGVCK 1245  
Qy 1321 -----GYNERADVPVSGVTHLNGNGRSYDPTLMRFTPTDLS--PRGAGINPYSY 1369  
Db 1246 EIGTVNBERPFGEOHDDERGLILRARY--YDPSVGRFTTKOVIGRRVTTGOSINRIVY 1303  
Qy 1370 CLGDPINRSDPSGH 1383  
Db 1304 TTMNPNVNLVDLTGY 1317  
  
RESULT 9  
ABG31849  
ID ABG31849 standard; protein; 2334 AA.  
XX  
AC ABG31849;  
XX  
DT 05-NOV-2002 (first entry)  
XX  
DE Human kinase, MEK1.  
XX  
KW Extracellular signal regulated kinase; hyperalgesia; surgery;  
KW opioid withdrawal; pain sensitization; analgesic; chronic pain; ERK;  
KW MEK1; human; enzyme.  
XX  
OS Homo sapiens.  
XX  
PN W0200258687-A2.  
XX  
PD 01-AUG-2002.  
XX  
PE 25-JAN-2002; 2002MO-US002128.  
XX  
PR 25-JAN-2001; 2001US-0264336P.  
XX  
PA (TEXA ) UNIV TEXAS SYSTEM.  
XX  
PI Gutstein HB;  
XX  
DR WPI; 2002-608414/65.  
XX  
DR N-PSDB; ABK30804.  
XX  
PT Reducing or reversing tolerance, physical dependence, hyperalgesia,  
PT withdrawal symptoms, or pain sensitization in patients on analgesics for  
PT chronic pain, comprises inhibition of the extracellular signal-regulated  
PT kinase (ERK).  
XX  
PS Disclosure; Page 156-161; 163pp; English.  
XX  
XX The invention relates to a method of reducing or reversing tolerance,  
CC reducing the risk of physical dependence or hyperalgesia, reducing the  
CC symptoms of opioid withdrawal or inhibiting pain sensitization in a  
CC patient taking analgesics. The method comprises administering an  
CC analgesic and an extracellular signal-regulated kinase (ERK) inhibitor  
CC comprised in a formulation to reduce or reverse tolerance, risk of  
CC physical dependence, hyperalgesia, symptoms of opioid withdrawal, or  
CC inhibiting pain sensitization in patients taking analgesics for chronic  
CC pain or those undergoing surgery. The present sequence represents the  
CC amino acid sequence of human MEK1 (not defined)  
XX  
SQ Sequence 2334 AA;  
  
Query Match 4.9%; Score 432; DB 5; Length 2334;  
Best Local Similarity 20.4%; Pred. No. 5.2e-20;  
Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;



```

QY 146 TGFNNAFDLKVPRKLL-----NPAGHAIYIDMN-----FEATQPLNRITVDLQGH-- 192
DB 1005 ---QINNAFYFKKGGKLOKVVYDGHNNATVYVNDNQULTAITDAGRKLLTTPDS--NGHVT 1060
QY 193 -----DIPLANLEY--QGLI-----KTLITLFPG 214
DB 1061 SINGPKKAVTSYENDLKKVTDQGTVSVDYDSGRLVKQYGSANSTBAKVFTEY-- 1118
QY 215 QKSGYRTFLRLKQLNSIHNF-----LGNENPLTWSFGTTPPGKNGILQGMTSM 267
DB 1119 QVSGHRLERKAINAKKEVYVSYADAKKTLTMTQPNGRKQVGYNEANP-----IQVID 1172
QY 268 APGLKKTVA--VSNNN-----QGHFPOSANLPLVYTLKQVPGAG 309
DB 1173 DAGSLKTTNTKEGNNVEDVDPNVGTGKATESYQYDKGN-----VTSVDAVGT- 1225
QY 310 QPAIOAEVSYTHSHVYGGSGNGIMNNKLDNLGYLMEYNYGSTERRYKDEGHDIIVR1 369
DB 1226 -----ETVEYNKNDV-----TKMKDEGNVTDIAVDGLDAVSETDQSGKSSAAV 1271
QY 370 ERTYNNYHLLTSECKQONGYITQTEYAVAIIGHNPDQSPSQPOLPKTETWRSADNSY 429
DB 1272 YDKXGNOIQSSKDLASSTNLIK-----DGSFEAQKSGMWLTASKO----- 1311
QY 430 RSEITETTPESGPNPLKVIKKDKTKKILSPS-----THMEYPPAGVNDCEPPEYFTR 485
DB 1312 RRRKIS-VIAPKSG-----VLSSGKALEVLSTOSTSAGTDHGISATQVLEBPNTTYLSC 1365
QY 486 FVKKII---QTPYSEFKDDEPEKIYQ---RYSLIGSOHVTLKIEBRHYSATOLLNS-- 537
DB 1366 KITDMLAKSAYRVIDARDQKRIQWINEYSLAKNDWT-----KROITFTTPANAGK 1421
QY 538 ---TLFOYNTDKSELGR-LKQTECTKENGKTYSVNHKFTYTKODDTLOOS--HSITTHD 592
DB 1422 AVVYMEVDHDKDKGKAMFEDEVQLEKEGEVSSSYNPQNSSFTSATENNVVSGASVDSSE 1481
QY 593 NFIHNSQVRSRYTRGLFSDTDFKDIYTKMSYDLGSLTFT-----L 635
DB 1482 GFNDVSVLKAKARTSASQAGSVTKQTVVLQGSANDKPVYTLTLGMSKASVKTDEKDYSL 1541
QY 636 NSGTPYANTLTLYDEL---NNLQDDNRPPFVI----- 664
DB 1542 QANVTYADSGTGYNAKFPBGTOEMRAAVIIPKTRINKVVISILFQKSGATGVWDDI 1601
QY 665 -----TTDVNGNQLRNEFDAGRHVSQCLKSDGDKFYTLITHOYDQGHHT 714
DB 1602 RLIEGSLTKSTYDSNGNYTKEDDELGVATS---TDYDETKG---KTSETDAKGEKTY 1654
QY 715 STY--SPYLTNGRQOTDPDKVHLSMSKSYDNWG-QIANT-----HHSYGVSEKI-- 760
DB 1655 YTYDQAQLTNMTLSNGSTSLH---SYDKENGEVSKTRIRAGADQTYKKEYDVMGLVK 1709
QY 761 TVDPITLTATKOLQSNENNVOI---GKEVTTYTPSQOPIQITLFDAGHLQSCHTLTRD 816
DB 1710 TTPBLGAVNLASBYDANSNLTKTISPNGEV-----SLSYD 1744
QY 817 GMDRVKRETDALGOCTIYQYDNVNRVITQLPDGTIVNRKXAPSTDTLLTIDRVNGISL 876
DB 1745 GTDRVKSXSYNGTEKYLFTYDKNGN-----ETSVNKEGN-----TT 1781
QY 877 GQOTFDGLSLTQSGDGRWATYYSAGNDQCESTVTPBGOPIHYOYEBLDAVLAQVA 936
DB 1782 KKRTPDKNRLTELTDRGSGQWVTPBDSKLKPF-----SWHH-----G 1821
QY 937 SNEITQOFSYNPVYTGALLKAVAGOSLPIYYPBGRMKMENIMDKMSYMLTLRGLENG 996
DB 1822 DQKGTNQPTYN-----KLMQMTKMSSTSSYSDYEN- 1854
QY 997 YTDITGIIQKISRDTHERVTOIKDSSIKTTLANDLNRHIGSQVTDLATGHMLTTTVEFD 1056
DB 1855 -----GNVQ-----TFITNGGCTSPSYDERMLVSLIHGDKNGDILITESYEV- 1898

```

```

QY 1057 GLNREIGRKLCSSGHTLIDQSSWLTQOLANRIVLNGVLOREQYSDSRNRLNYKC 1116
DB 1899 -----DANGRRITINS-----ASGRV---QYEGKLNQLVK--- 1927
QY 1117 DGABCEPTDKYGHISIVTQNEFTYDIYGNITACHTTFADGTED--HATKFPANPTDPCQITVEY 1174
DB 1928 -----ETHEDGVYI---EYTYDGFGRKRTV--TTIKDSSSKTVASFNIMN-----OLTKV 1973
QY 1175 H-----HTHP--DMPDN-----RLKDYKGRVINITDNG 1203
DB 1974 NDESISYDKNGNFTSDGKFTYTWDAEDNLTAVYKGEDKPFATYKYDEKGNRIQKTVN-G 2032
QY 1204 NTFENFTYDTLGRLONGQGSVYGYDPLNRLVSOQTDITLDEELY--YRETMVLVNEVRNEMIR 1262
DB 2033 KATNYFYDG-----DSLNLVLETADNNVTKSYTGD-----SQGLLS 2070
QY 1263 LARTGETTIAQORASHVLLTGTDSQOSVILTSDKONLSQEAAYSAYK--HKSTANDA--S 1318
DB 2071 YTENGKKVFYHYAHDIIAISDSTGTV-----AKQYDAMGNPTTEASDEVKDN 2122
QY 1319 ILGYNERPADPVSGVTHLNGRYSYDPTLMRHT--PDSLSPPGAGINPYSCLDGPRIN 1376
DB 2123 RYRYAGYQYDEETGLYLLMAY--YEPNRGVFLSDPDPSDSDSLDQNGYAVAGNNPVM 2180
QY 1377 RSDPSGHLSSQAWTQIGMGIAGLLTLTATGMAIAAG---GIAAIASTSTTALAFA 1432
DB 2181 NVDPDGH--W-VMLVYNAGFA-----AYDGYAYYSGKMGKMAAANAANPBGKIFKG 2231
QY 1433 LSVTSDITSIVSGALDASPKASSILGVWSMGW---GAAGLAEASIKGGTKLATHLGAF 1488
DB 2232 ASRAYKFTK-----KAVKITGHTRHGLNQSIGRNG-----GRGVNLRALKNA- 2273
QY 1489 AEDGENALAKSTSESSRIKGVTRSLDR--LYRNEBGQYIKDHSRGYDNPFGKGEQAI 1546
DB 2274 -----VRSPPKCVIKOPNGATKYGVKATVVLNRKGVITAYG-----SS 2312
QY 1547 LVHGDQDGLYHTEGNK 1563
DB 2313 RAKSGHVFHTHKGKMK 2329

RESULT 10
ABU18641
ID ABU18641 standard; protein; 2234 AA.
XX AC ABU18641;
XX 19-JUN-2003 (first entry)
DE Protein encoded by Prokaryotic essential gene #4168.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Bacillus anthracis.
XX PN W0200277183-A2.
XX 03-OCT-2002.
XX PD 21-MAR-2002; 2002MO-US009107.
XX PF 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KJ, Zysek JD,
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
XX WPI; 2003-029926/02.

```

DR N-PSDB; ACA22511.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 46565; 1766bp; English.  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 623 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC the gene product; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC identified for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-regulated gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2234 AA;  
Query Match 4.3%; Score 386; DB 6; Length 2234;  
Best Local Similarity 18.7%; Pred. No. 8.3e-17;  
Matches 339; Conservative 236; Mismatches 633; Indels 608; Gaps 76;  
QY 21 PTOANNFNSVNSGVDR-----TGLYNIDITLGHVY-----NGUL----- 57  
DB 748 YSKVGHASEISISHEFPFRPSODSGFIMGFDWASIPVINGKVNATNGFMNSEKDITLSG 807  
QY 58 -GPTPLPLTSLSPKNTDIGNFGIFNFGLSV---VDRKNSLLSLG-----TGEN- 102  
DB 808 RGPVSVERTNSGKVGLEFSTGWSGLEERVMADGNGNLLITDGNATITFTTGDK 867  
QY 103 -----YKVIETDKTVKLOQKQDNLRFKDKLKENCYRIHKSGLIEVL 145  
DB 868 YOAPGTIVLEIKOVGSGYEIKDKQTV-----TFKSGDAQGR 905  
QY 146 TGFNNNAADLAKP-----KLLNPRAGAIYIDMFEATQPLRLNIYDDLDHDIPL 196  
DB 906 IETYDKRGNTTVEYDGASRLSKVKNASGKELV--GDGNNKKAAYV-----GPDKXT 959  
QY 197 INLEYGGLIKITLTFPGQ--KEGYRTELRLPNROLNSIHNSFSLGENPLTWSFGY----- 250  
DB 960 ITFNTDGLLVSTTPEGKVKYCYD-----NGULTSITDPOHTDAPRYKTSYAYEBDR 1013  
QY 251 -----TPIGKNGIL-----GQWITSMAPGLKLETYNVSNNGHHFPOSANLPLPYVTL 301  
DB 1014 LVKATVDPGKATTLAANTGSKEVTLTPNGKRTVYTYND-----AGRPV----- 1057  
QY 302 MKQVGAQGPALQAEVSYSHNYVG-----GSGNGIWNKKDLNLYGLMTEYN- 348  
DB 1058 -KTVEDVGLNLITTSYEYNNANLVKTTTPKQNETATYDNGNNTSVTDEMGTEKEFEYNK 1116

QY 349 -YGSIESRRYKXE-----GHDQIVRIERTYNNYHLTSECKQONGYIQTETAYAYAI 400  
DB 1117 DNGIKATDNEDKRTIYAYVANGTVEGQTDGAN-----TSSVHHHQYGNPIETSELS 1171  
QY 401 IGHNPDSPQPOLPKTKETWRSADNSYRSEITETTFDSGNPLTVIKDKTKQTKIISP 460  
DB 1172 AGGNLIGNPS-FEM--NGTEKVVAVDPTNNSGISIS-----KDATPAPGLGGESESLKITYTK 1223  
QY 461 STH--MEYYPAGVNDCEPBPGFTRFYVKII-----QTPYDESPDDP 503  
DB 1224 ATNNDKQIYALIQETLEPMTTYTLTSGMXTDLVNGAFAFRVOSLBNAGACIDGMDTR 1283  
QY 504 EKPIQYVYSLIGSQSHVTLKIEBRYSATOLNLSLTFQYNTDKSELRL-LKQTECTGGE 562  
DB 1284 HNKVQGSIDWVNRQ--VTFPTTEQ---TRKVKYLVENGSGATSSGAWPDKIQLEKGE 1337  
QY 563 NGKTVSVHKFTYTKQ--DITLQO-SHSITTH-----DNFTIHSQVRSRYTGLRFS 611  
DB 1338 VSSSFNFVLSNSSFENNPPDFVPQWVWSCQHCERNVDSDSPFGHSSIVMER-----S 1391  
QY 612 DTDKDIIVTQMSYKJGRLLTRTLNSGTPYANTLTVDYELNLQDDNRPPVITTTDVNG 671  
DB 1392 EYGENDI-----GYRNRVILNOKKATVTTLTAMSKSENVND----- 1428  
QY 672 NQLENEFDGGRHVSOCLKDSDEGKFYTIHTQOYDEQGRHITSYSDYLTNGRQOTDP- 730  
DB 1429 -----APDKLSKOYAVLAERTYDQ-----TVVNYTSPSGNTDM 1464  
QY 731 -----DKVHLSMSKSYDNWQOLANTHWSGVSEKITVDPITLTATQLOGNS 777  
DB 1465 NRSAAVIPAACKPLOKIKIFILFRKNNG-----KWPFDIRLLEGNALIKNE 1511  
QY 778 NNVTGKEVNTYTPSQPQIOTLPEBAGHLOSCHTLTRDGDWRVKEKETAIGQCTIYOYD 837  
DB 1512 YD-NDGNVAVTY-----DEBQGR--NTFTYDASGNGKSETDEGNTKLDYN 1555  
QY 838 NYNKVIQITLPDGTIVNRKAPFSTDTLITDIRVNGISLQOTPEDGSLRLOSQDGGRW 897  
DB 1556 KDNLLTRKVTLLKNGTSVAVRYD----- 1576  
QY 898 AYTYSAQNDQCEPVTITPDGQFIHYQOPELDAVLOVASNETIQGSPYNTGALLKAV 957  
DB 1577 ---HNNTTEKSYWFGKQTHREYDVVNKNVTYIDALNRREN--TYDENANKIKITKM 1631  
QY 958 AEGSLTPYYPGSLKMEMENINDMKMSYLTTLGLENGYDTLTGTLQIKSRDTHGRVTO 1017  
DB 1632 PNGSILSVDYDADRIVGEK-----RNGKDSFT-----FERDONGVYTK 1670  
QY 1018 IKD--SSIKTTLVNDLNRHIGSQVTDLATGMLTTVFPDGLNREIGRKLCDSSGHTLD 1075  
DB 1671 VKDLVNGVERTKYDKADR-----VTS-AT-----DSRGKID 1702  
QY 1076 IQGSMV-----KTQOLANRIYKLVNGVLQRTBOYSYDSRRRLNO--YKCGAACPDPKY 1126  
DB 1703 ---WAHYDKANSKTEBKKEQTVQGY--TNKVSQ-YVTLIDONIRVTTGSSQ----- 1748  
QY 1127 GHSIVTQNFYDIGNITACHTTPADGTEDHATFKF--ANP-----TDFCOLTEVANT 1177  
DB 1749 ---TRFDYDDQNV---RTTYAANGSGSTINYQANKIKOLVVGTSISILSEXYE 1799  
QY 1178 HPDMPDNIRLKYDRAG-----RVINITNHNQNTENFTYDTLGR-----LQ 1217  
DB 1800 YDQSGNRTKIKHGAAGGVETETNFVYDPIQLNLBEVLPNGTYSKYTDGFGNRTSVKVI 1859  
QY 1218 NGQGS---VGYDPLANLVSGKDTDLDCEL----- 1244  
DB 1860 NGKETSIAATFNGNQLVKGNESLTYDVNGNRTSDQKRYTWNEDQIVALTKQGENN 1919  
QY 1245 ---YRETMVLVNEVNGEMIRLRTGETI-----IAQ 1274  
DB 1920 AFATYTKDEDENRRIEKVNGGVTRYFYDGDGINPLYETDNGVTLRQYVYSGADGARLANK 1979

QY 1275 RASKVLLTGTDSQOQVI--LTSKONTQSEAYSAVK-----HKSTANDASILGNGERA 1327  
DB 1980 AAGCTLYHNPRDQVAMNODKEVATYEDYMGVNLSDTGIAAD--NPEYIAYMY 2038  
QY 1328 DPVSGVTHLGNGYKSYDPTLMRFPTD-----SLSPFGAG-----GINPYSCIGDPINR 1377  
DB 2039 DKEIGMYLL-----IARYNPEHGVFLSYDPDGEDDPTVMNGTYADNNPVM 2088  
QY 1378 SDPSGHSKQAMTIGMGAGLLTITATGMAIAAAGIAAATSTTALAFALSVTS 1437  
DB 2089 TDPGKAM-----LVPVYIAGAMVAFGAKYAI-----RYGAKYQKAKVKS 2131  
QY 1438 DITSVGALEDAEPKASILGWSMGGAAGLAESAIGGTGKATATGAPAEDEGVAL 1497  
DB 2132 -----GNDY-----GKKVAKSGMNGKSTIAQKIPRIHKVR---I 2163  
QY 1498 KSTSESRIRKMGVTRSLDREIVRNEGOVTKDSRGYTDNFMGKEQALIVHGDKDFLY 1557  
DB 2164 KGDNDKKGKYGWVYISTTK-----KTKGRYS-----SFEHTPHNGHY 2203  
QY 1558 HTEGNKNG-KGPYTR 1572  
DB 2204 HLOKNKYSKYQKWMR 2219

## RESULT 11

ADCC1365 ID ADCC1365 standard; protein; 1400 AA.

AC ADCC1365;

DT 04-DEC-2003 (first entry)

DE Enterohaemorrhagic E. coli O157:H7-specific protein SEQ ID NO: 1410.

XX enterohaemorrhagic; anti-bacterial.

OS Escherichia coli; O157:H7.

FN JP2002355074-A.

PD 10-DEC-2002.

PF 24-JAN-2002; 2002JP-00015959.

PR 24-JAN-2001; 2001JP-00112010.

PA (UYTS-) UNIV TSUKUBA.

DR WPI; 2003-451640/43.

XX Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule  
PT and a polypeptide and its use; a polypeptide, a vector and a host cell.

PS Claim 3; SEQ ID NO 1410; 2067bp; Japanese.

XX The invention relates to a novel enterohaemorrhagic Escherichia coli  
CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention  
CC has anti-bacterial activity. The polypeptide can be used in detection  
CC and/or treatment of O157:H7 infection. The nucleotide sequence of the  
CC genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present  
CC sequence represents an E. coli O157:H7-specific polypeptide of the  
CC invention.

XX Sequence 1400 AA;

QY Query Match 4.1%; Score 365; DB 7; Length 1400;

DB Best Local Similarity 20.9%; Pred. No. 1.2e-15; Indels 456; Gaps 65;

QY 418 KTEEM-BSADNSYRSSEITETTFDESGNPLTKVI-----XDKTKOKIISPSTMEY 466  
DB 218 RTLYREBAAGDLAGEITGV-TDAGRGRFRLVLTQAGRAEARKQHTASLSPDT----- 272

QY 467 YPPAGEVDCPRPYPGTFPRVKKIIOTPYDSEKDDPEKIOYRSLISQSHVTLKIEE 526  
DB 273 -----PAPLSDSAPDPTLPGETY-----GPDKIR----- 297  
QY 527 RHYSATOLLNSTLFOYNTDKSELGRLLKQTECTGNGKTSYV-----VHKFTYKOD 579  
DB 298 --LSAVWLTHDPAPVPSLPLPAPLARY-----TYTAGELLAAYDSSNTQVRAFTVDAQH 349  
QY 580 DTLAQSHSITHDNFTIHSQVRSRYTGRLPSTDTKDVTQMSYDKLGRLLRTTLNSGT 639  
DB 350 PGRMVAN-----RYAGR-----PEMRRYRYDTGIVVQQLNAGL 383  
QY 640 PYANTLTYDELANLDDNRPPVITTTD--VNGNQLBNERDGAHRYSQCLDSDGKXF 698  
DB 384 SY--RYOYE-----QDR-----ITVDSLNREBVLHTEGAG--LKRVKKEADG-- 425  
QY 699 YTIHTQOYDEQGHHTSTYSDYLTNGR--QOTPDVKHLMSKSYDNMQIANTHMSYGV 756  
DB 426 -----SVTHSGYDAGRLLTQTD-----AAGRRTYGL 453  
QY 757 SEKIYVDPITLTAIKOLQSNNSNVQGEVYTYTPSQOPIQITLFDAGHLQ--CHT 812  
DB 454 N--VVSGLDITDITPDGRRETFYNDGNQLTAVSPDGLSRBAYPEPRLVSETRCGD 511  
QY 813 LTRDGMRVKKE-----TDALGCTIYQYDNVRVITLPDGTIVNRKAPSTDTLIT 867  
DB 512 VIRYADNPHESLPATTTDASTTROMTSRYQLAFDTCSGYQRYRDEDRPQMTAVH 571  
QY 868 DIRVNGISLQOTPDGLSRLOSQDG--GRVATYASANDQCPSTVTPDGPFIHYQOP 926  
DB 572 --REBGIS--RYRYVYDNRGLTSVKDAQGHETREYNAAGL--TAVITPPGNSEYQ-- 624  
QY 927 ELDDA--VLQVANSBITQOFSYNPVYTGALLKAVASQSLPTIYPSGRLMENINDMK 983  
DB 625 ---DAMGKAVSTQGGILTRSMEDY-LAGRITTLTNGRSRSEFTYDA----- 667  
QY 984 MSYLMTLRGLENGTYDITGIIKISPDTHGRVYQIQQSSIKITLNYDNLNRHG----- 1037  
DB 668 LDRIVQGRFPD-----GRTORHYDITGKLTQSEDEGLVTMMHDESRDLTRHYNGE 720  
QY 1038 -----SOVTLATAGMLTTVAFDGLNREIGRKLCDSSGHTLDI-----QQS 1079  
DB 721 PAKOMQYDEHGMWLTEISHSEGHVAVHYGDKGLAGSRQVNHNPETELLMQHETEH 780  
QY 1080 WLKTQOLANRIVKLVNGVLRQTEQSYDS-----RNRL-----N 1112  
DB 781 AYNQGLANRVTP--DSLPRVEMWLTGSGYLAKMLGTEPLVEFTDRLRETVRSFGNN 838  
QY 1113 QYKCDAGECTDKKH-----STV-----TONFTDIYGNIT 1144  
DB 839 AVELSTYTTA--GHIQSQLNSQYDRDYDWNNDGLVIRISGPRTWEGYSANRLE 895  
QY 1145 ACHTTFADGTEDEATFEKANTDPC--QLTEVHHHTPD-----MPDNIRL-----KY 1189  
DB 896 SVRTIASD-----LDIRIPATPDAGNRLPD--PELHPDSTLTAMPDN--RAEDAHVYRH 948  
QY 1190 DKARVINITD-----NHGNTENFTYDLSRL-----QNGQSV--YGYDPL-- 1229  
DB 949 DVGRLTEKTRIPAGVIRTDDBRTHHYDSQHRLVFTYRIQHGEPVLSRYLDPLGR 1008  
QY 1230 -----NRVLSQKTPYDLCELYR----- 1247  
DB 1009 RMAKRVRRERDLTGWMSLSRKPEVTWYMGDGRLLTVQDTTTRIQVYPPGFTPLIRV 1068  
QY 1248 -----ETMLVNEVRNG-----EMIRLRTGETIITAOQAS----- 1277  
DB 1069 ETENGEREKAKQRSIAETTLQOESENGHVFPALVRLDLRLEERLRARVSESAWML 1128  
QY 1278 -----KVLITGTDSC--QSVILTSKQNTSQEAYSAVGRKST 1313  
DB 1129 AAGCLTVEQALARQVEPEYTPARKVHFYHCDHRLPLALISDQNTAMRGEYDEMGQLTNE 1188





```

QY 1170 Q-----LLEHHHTHPMPDNIRLKDYKAGRVINIDNNGTENTFYDTL 1213
DB 578 QGETAQAAGSVPFENRITRYGLH-----YRDEYGVVEKGRNG-TQHYRMDAE 628
QY 1214 GRL-----ONGGGSVYG--YDPLNRLVSOQKTDTLDCLELYRRTMLVNEVNGEMRLLR 1265
DB 629 HRLLEVAVTGGTYRRYGVYDAPGRVVEKHEDBAEKFPNRTTFLMDGRRLAQECRLGR 688
QY 1266 TGETTIAAQR-----ASKYLLTSTD-SQGSVILTSKONLSQEA-YSAIY 1307
DB 689 SSSLVIYSDRGSHPEPLARVRAAPGEADEVLVYHTDVGAPBEETDGGNIWEAGYQW 748
QY 1308 GK---HKSTANDASILGYNBERADPVSGVTHLNGVRSYDPTLMRFPTPSLSPFG-AGG 1363
DB 749 GNLTHEKETRPVQONLRFQSQYLDREYGLH--NLVRFYDIDGKFLTSGD---PIGLAGG 803
QY 1364 INPYSYCLGDPINRSDPSG 1382
DB 804 INLYQVA-PNPLSYIDPLG 821

RESULT 14
ABU19676 ID ABU19676 standard; protein; 1515 AA.
XX
AC ABU19676;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by prokaryotic essential gene #5203.
XX
KM Antisense: prokaryotic essential gene; cell proliferation; drug design.
XX
OS Borrelia cepacia.
XX
PN W0200277183-A2.
XX
PD 03-OCT-2002.
XX
PE 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342823P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI WPI; 2003-023926/02.
XX
DR N-PSDB; ACA23546.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 47600; 1766bp; English.
XX
XX
The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation or the activity of a gene in an operon required for
proliferation; (7) identifying a compound that influences the activity of

```

```

CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strings is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1515 AA;

```

```

Query Match 3.9%; Score 350; DB 6; Length 1515;
Best Local Similarity 22.1%; Pred. No. 1.5e-14;
Matches 232; Conservative 126; Mismatches 314; Indels 380; Gaps 59;

```

```

QY 481 YGFTRPYKAKIQTPIYSEFEDDEKFIQYRYSILGSHHTLKIIEHRYATQNLSTLF 540
DB 543 FQFGQFYAYADENGMWTQWEDTDQTVRYRYDTAG-----RVV-----TGTRQ 586
QY 541 QYNTDSESLRLKQTECTGKNGKTYSVVAKFTYTKQDPTLQOSH-----ITTHDNFTI 596
DB 587 GYHT-----GRFYEAQCT-----RVIDVGGMTVAANDGGLVTAERDPLGHCTYSMBL 636
QY 597 HRSQVR-----SRV--TGRIFS-----DITDKOIVT-----QMSY 624
DB 637 GRIMARIDPLGRRTDYDERGQLTSVBSGRTVDYDDDEQLNARLPNGSTIYLBV 696
QY 625 DKGLRLTRTLNGSTPYANLTLDY-----EL-----NNLODDRPVPTITTYD-- 669
DB 697 DHDRLIART-----EPDGNKTYRYGREGELRVAGQDRETRLDYDR-----LTLIDIEL 748
QY 670 -NGNQLNEFPDGAGRHVSQCLKSDGDKFYTTHTQOYDEGSRHHTSYSDYLTNGR--- 725
DB 749 PTGARFRPKIDALGR-----LLEETSPDG-----HYTRKYD-----YADGRANPGLL 790
QY 726 -QOTDPPKVLHLSMSKSYDNWGOIANTHWSYGUSEKITV---DPTTLTATKOLQGSNNVQ 781
DB 791 SAVTRPD-----GSVSRARYN---SESLFVEMIDPLGRYTOR----- 824
QY 782 TGKEVTTYTSQQPIQITLFDGSH-----LQSGHTLTRDGMRYRK 823
DB 825 -----TYGPPD--LTLASIDAGAHATREYDHATRLTKVINALGETTYTRYDAAGRIAA 876
QY 824 ETDIAIGQCTIYQYDNVNRVITLTPDGTIVNRKYAPSTDTLLDIRVNGISLQO--QTF 881
DB 877 EIDMGGAITYDRARAVERLLTKLPDG-----GQWRYTY 910
QY 882 DGLSRLTQSGDGRVMAVYYSAGNDGCPSTVITPDQPIHYQVPELDDAVLQVANSBIT 941
DB 911 DASDRLIEIDAGDVKCLAYRYDASG-----RLASAVVQGEHTHTVT 949
QY 942 QQFSYNTVTGALLKAVAGGSLTFIYPSGLKKNENINDMKMSYVWLTIRLNGYDTLT 1001
DB 950 -RFAYYDR-NGLLIGEDGHGELLRHVYADGQRR--RMTPRRETTYAY----- 993
QY 1002 GTIQISRDTFGRTVQIKDSIKTTLNVYDDLNRHIGSQVDTL-----ATG---HMLT 1050
DB 994 -----DVSGLATVGG-----QLTIRRDGGRRIEIGBAGFVAQQQYDALGRIRROJAG 1041
QY 1051 TTVFPGINREIGRKLCDSSGHTLIDIOQSWLKTQOLANRYKLVKNGVLTORTQYSDSKNR 1110

```

Db 1042 PAVAFDALQADPARAL-----EQLTRQV-----YHYDAAG 1072  
 Qy 1111 LNOYKDGACRPT---DKGHSIV-----TONFTDIYGNITACHTTADGEGEDATP 1160  
 Db 1073 LERVVD-TGADTLTYQDERGQIICAESLQPSSEFRYDAWNI-AAHGORA----- 1121  
 Qy 1161 KPAFPTDPCOLTEVHHTH-----PDMPNIRLKYDKAGRVINIT-DNHG---NTENFTYDT 1212  
 Db 1122 -----PVD-----AHHRRGGLPEQGVYARYKDKADARRTIEKVEQGVAPKTYQWTWDG 1171  
 Qy 1213 LGFL-----ONGGSGVYGYDPLNRLVSOQKTDLDCELYRET-----MLVN---EV 1255  
 Db 1172 LNLVLVVTPEERGVMWA-YRYDAEFNRRIEQ-----QVGREGTYKFLMDGPMALAEHWIEQ 1224  
 Qy 1256 RNEEMIRLRTGETI-----IAQORASRV-LITGTSQGSVILTSKONLSQE 1302  
 Db 1225 RDG-----TTGGVVVWHIEPGSFLPLAQSTBDGLFPLIDQIGRPVTEDEGCRPWKA 1278  
 Qy 1303 AYAGVAK---HKSTANDA-----SILGNGERADPVSGVTHLNGVSYDPTLMRFH 1351  
 Db 1279 AYSLMGLPLPVKRPADADGATSIDTTLRFSGQMDDEFGVYNNLRY---YPPDSQYL 1336  
 Qy 1352 TPDLSLPPG-AGGIMPYSYCLGDPINRSDPSG 1382  
 Db 1337 SAD---PIGLGAGARTQAY-VHDPSPQWIDPLG 1364

RESULT 15  
 ABUL5135  
 ID ABUL5135 standard; protein; 1397 AA.  
 AC ABUL5135;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by prokaryotic essential gene #662.  
 XX  
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Escherichia coli.  
 XX  
 PN W0200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002MO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselebeck R, Ohlsen KU, Zykkind JW;  
 PI Wall D, Treweek JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA19005.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS claim 25; SEQ ID NO 43059; 1766bp; English.  
 CC  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPD at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
 SQ Sequence 1397 AA;  
 Query Match 3.9%; Score 342; DB 6; Length 1397;  
 Best Local Similarity 21.4%; Pred. No. 4,8e-14;  
 Matches 256; Conservative 157; Mismatches 382; Indels 404; Gaps 63;

Qy 418 KTBTM-RSADNYSRSITETTPESGNPLTKV-----KDKTKQKIS-----PS----- 461  
 Db 218 RTQTFHREAGERSGHTGV-TDAGRHFELVLTQORAEARQOASIGTEPSAPDPT 276  
 Qy 462 -----THEEYYPAGEVNCPEP---YGF-----RFVK 488  
 Db 277 LPGTYERGNDGIRLSAVALTHDPERP-----ENLPAALPVRYGWTFRGELAAYDRSNT 331  
 Qy 489 KIIQTPYDSBFK-----DDEKFTQYRYSLSGSHVTLKIEBRHYSATQLNSTL 539  
 Db 332 QVRSFTYDDKYRGMVAHHRHTGRPE--ICRYD--SDGVTBQLNPAGLSY----- 379  
 Qy 540 FQYNTKSELGRLKQTEC--TKGNGKTYSVYHKTYYKODDTLOQSHSITTHDNFTIH 597  
 Db 380 YQYEKDRITITSLNREVLHTGEGG-LKRVVK-----EHAQGSYV 421  
 Qy 598 RSQVRSRYGRLEPSDPTDQIVQMSYDLGRLLTPTLNSGTPYANTLTGYDELMNLQDD 657  
 Db 422 QSQFDA--VGRLEAQTDAAGRITTEYSPDYVGTGLTRIT--TPDGRASAFY----- 468  
 Qy 658 NRPFFVITTTDVNGQLRNEFPDAGRHVSQCLKDS-----GDKEFYTHITQYDEGGRH 712  
 Db 469 NHHSQLTSATGPGGLIRREYDEWGRLLIQTAPDGIITRYRYNPHSDELPCATEDATGSR 528  
 Qy 713 HTSYSDYVLTNGRQOTDPKCHLSMSKSYDNWQQLANTMSYGVSEKITYDPI-TLPAIK 771  
 Db 529 KWTWMSRY--GQLSLFTDCSGGYVTRYDHRFQVAVVAREEGLSQRAYVDSRGQILAVK 585  
 Qy 772 QLOSNNNVOTGKEVVTTPSQOPITLFPDEAGHQSCHTLTRDQMDRVRKXETDLAGO- 830  
 Db 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG-SRNGTQYDMWGRK 623  
 Qy 831 -CTI-----YQDYNRNVIOITLPDGTIVNRKVAFPSTDTLITDIRVNGISLAGQTF- 881  
 Db 624 ICTTQGLRSMHYDAAGVIRLTSNGS-----HTTFRYVDVLDRLIQTGFGFRGORYH 678  
 Qy 882 -DGLSLTQSGQGRWATYTSAGNDQCSYVITTPGQITRIHQYQELDVAIVQVANSNEI 940  
 Db 679 HDLTGLIRSEDEGLV-----THMHTYD-EADRLTRTYNGEFT 714  
 Qy 941 TQGFSTNPVTGAL--LKAVAGQSLTPRY-YPS-GRLLKMEKI-----NDM---KQMSY 986



Db	715	AERQOYBE--RCWMLTDISHISBGHVVYVNYGVDSKGRLASHELVYNNHQTNELMONGHTRH	773
QY	987	LMTLRLGNGYTDLTGTIOKISRDTGCRVTOIKDSSIKTTLNYDDLNRHIGSQYTDLATG	1046
Db	774	AYNAQGLAN--RCIPDLSLPAVEMWLTYG-----SGWLSGKMLG	808
QY	1047	HMLTTAYEE--DGINREIGRLCDSGGHTLIDQOSWLTQOLANRYLVNLORTQESV	1104
Db	809	D--TPLYEYTDRLHREYLR-----SBRXYELTAYTPAGLOQO--HNLBLS--DRDYT	858
QY	1105	YDSNRRLNQYKCDGACBCEPTKYGHSIVYQNFYTDYIGNITACHTTEAD-----GTEDHA	1158
Db	859	WINDGELLR-----ISSPRO-----TRSYSVYTGRLTGVTHTTAALDIRIPYTEDPA	906
QY	1159	TFKXANPDPBQCLTEVNHHTPD-----MPDN-----IRKXOKAGVINITD-----	1200
Db	907	GNRLDP-----ELHPDSALSMWPDNRILADAHLYLYRDNHGLTERTKTDILPEGV	956
QY	1201	---NHGNTENFYTDYTLGRLONGQGSVYG-----YDPLNRLVSQKTDYLDCEL-----	1244
Db	957	IRTDERHARHAYOSQHRLLVHYTRQVAPLVEKRYLYLDPGRYAKAKVMBRERDLTGM	1016
QY	1245	---Y----KETLVNE-----VRNGEMRLRLR-----TSETIIAQORA-----	1276
Db	1017	SLSRKPQVYTWYGMDBDRLLTITONDMTRIQITYOPSSFTPLRLVETATGELAKTORSLAD	1076
QY	1277	---SKYLTGTSBQGSVILTSDKONLSOEK-----	1303
Db	1077	TLOQSGGEDGGSVVPRLVQMLDRLEBEILLA--DRSEBSRRWLASCGLTVAMQMSQM	1133
QY	1304	---YSAYGKHSSTANDAS-----ILGYNG	1322
Db	1134	DVVYTPARKIHLVHCDHRLPLALISTEGTTAMVAYEYDEWGMILLMBENPHOLOOLIRLPG	1193
QY	1325	BRADVSGVTHLNGCYRSDYPTLMEFPHRPDLSPEFG--AGGINPVYSCGDPIPNESDPFG	1382
Db	1194	QOYDEBSGLYX--NHRHYTDPLOGRTYTD--PGLGKGANWFYQPL--NPLSNIDPLG	1246



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 30, 2006, 09:42:59 ; Search time 37.5112 seconds

(without alignments)  
3687.335 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879

Sequence: 1 VYIKFLKLRRTMSDNNER.....PKIILGRTEKTVKPTFRP 1673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5 COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/6 COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/8 COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/PCUS.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/RB.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1638	18.4	1584	2 US-09-251-645-6	Sequence 6, Appl1
2	339	3.8	1426	2 US-09-492-709A-340	Sequence 340, App
3	331.5	3.7	1377	2 US-09-711-164-467	Sequence 467, App
4	317	3.6	998	2 US-09-252-991A-28424	Sequence 28424, A
5	316	3.5	1627	2 US-09-328-352-6604	Sequence 6604, Ap
6	306	3.4	1586	2 US-09-543-681A-5329	Sequence 5329, Ap
7	294	3.3	1439	2 US-09-543-681A-7560	Sequence 7560, Ap
8	292	3.3	1626	2 US-09-252-991A-23805	Sequence 23805, A
9	285.5	3.2	1596	2 US-09-328-352-5542	Sequence 5542, Ap
10	277.5	3.1	974	2 US-09-252-991A-23640	Sequence 23640, A
11	271.5	3.1	1665	2 US-09-543-681A-4476	Sequence 4476, Ap
12	238.5	2.7	1043	2 US-08-851-567B-61	Sequence 61, Appl
13	235.5	2.7	804	2 US-09-328-352-5545	Sequence 5545, Ap
14	223	2.5	1128	2 US-09-252-991A-31032	Sequence 31032, A
15	222.5	2.5	3290	2 US-09-328-352-5486	Sequence 5486, Ap
16	220.5	2.5	1183	1 US-08-447-031A-2	Sequence 2, Appl1
17	218.5	2.5	2123	2 US-08-968-685A-10	Sequence 10, Appl
18	214	2.4	2504	2 US-09-328-352-5821	Sequence 5821, Ap
19	212.5	2.4	954	2 US-09-251-645-12	Sequence 12, Appl
20	211.5	2.4	10182	2 US-09-134-001C-3119	Sequence 3119, Ap
21	207	2.3	2314	2 US-09-268-347-49	Sequence 49, Appl
22	202.5	2.3	2777	2 US-09-543-681A-6124	Sequence 6124, Ap
23	201	2.3	1739	2 US-09-540-236-7339	Sequence 3739, Ap
24	199.5	2.2	1052	2 US-09-499-203-2	Sequence 2, Appl1
25	198.5	2.2	1004	2 US-09-147-405B-15	Sequence 15, Appl
26	198.5	2.2	1004	2 US-09-268-347-30	Sequence 30, Appl
27	197.5	2.2	1726	2 US-09-700-227-2	Sequence 2, Appl1

28	194.5	2.2	5024	2 US-09-710-279-2964	Sequence 2964, Ap
29	190	2.1	2385	2 US-09-543-681A-6304	Sequence 6304, Ap
30	188.5	2.1	2283	2 US-10-172-502-4	Sequence 4, Appl1
31	188.5	2.1	2736	2 US-09-252-991A-30227	Sequence 30227, A
32	185.5	2.1	1166	2 US-09-200-650E-7	Sequence 7, Appl1
33	185.5	2.1	2383	2 US-09-492-709A-302	Sequence 302, App
34	184.5	2.1	2142	2 US-09-540-236-3459	Sequence 3459, Ap
35	184	2.1	3696	2 US-09-134-001C-5080	Sequence 5080, Ap
36	183	2.1	1222	2 US-09-206-942-37	Sequence 37, Appl
37	183	2.1	1222	2 US-10-193-764-37	Sequence 37, Appl
38	183	2.1	1228	2 US-09-206-942-34	Sequence 34, Appl
39	183	2.1	1228	2 US-10-193-764-34	Sequence 34, Appl
40	182.5	2.1	1565	2 US-08-851-567B-59	Sequence 59, Appl
41	182	2.0	1861	1 US-08-790-912-4	Sequence 4, Appl1
42	181.5	2.0	2315	2 US-09-543-681A-5434	Sequence 5434, Ap
43	181	2.0	1228	2 US-09-463-402-2	Sequence 2, Appl1
44	181	2.0	1228	2 US-09-889-572-2	Sequence 2, Appl1
45	181	2.0	1228	2 US-09-117-447-2	Sequence 2, Appl1

## ALIGNMENTS

```
RESULT 1
US-09-251-645-6
Sequence 6, Application US/09251645
Patent No. 6281413
GENERAL INFORMATION:
APPLICANT: Kramer, Vance C.
APPLICANT: Morgan, Michael K.
APPLICANT: Anderson, Arne R.
APPLICANT: Hart, Hope
APPLICANT: Warren, Gregory W.
APPLICANT: Dunn, Martha
APPLICANT: Chen, Jeng S.
TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR
FILE REFERENCE: CC01963/A
CURRENT APPLICATION NUMBER: US/09/251,645
CURRENT FILING DATE: 1999-02-17
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 6
LENGTH: 1584
TYPE: PRT
ORGANISM: Photorhabdus luminescens
US-09-251-645-6

Query Match      18.4% Score 1638; DB 2; Length 1584;
Best Local Similarity 31.0%; Pred. No. 3.1e-118;
Matches 493; Conservative 233; Mismatches 682; Indels 182; Gaps 54;

19 EFFTOANNFTSAAGGVDPRTGLYNIQITLGHIVGN--GNLAGPLPLTYSPLNKTIDIG 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5 DYSNAFNPESYINTGVDPRTGQYSANINITLRPNVNSBQT--LSLSFSLTLTANG 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
77 FGIGFNPGLSVDRKNSLSLSTGENYK--VLETDTYVLQCKKLDNIRFEKOLKENCY 133
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
63 FGIGFNPGLSVDRKNSLSLSTGENYK--VLETDTYVLQCKKLDNIRFEKOLKENCY 121
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 RIHKGSDIEVLGFNNMAFDLVPKCLNPAHAIYIDNPFATQPRNRIYDDLDGHD 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 YVNNKGIIHLRIKISS--DIKTVALEPDEADPLVNSR----- 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
194 IPIINLEYOGLKITILT--PFGQKEGYETLRL--NROLNSINNFSLGNEPPLTWSFGYT 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
163 FALSEIKRYRTGYTKLANSNN---CISVEPDDNNISAKAFVDRNDYLITVTPVD 219
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
252 PICK-----NGILGWIISMTAPGLKETTANSNNQGHFPQSANLPVLPY 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
220 ASGPIDSAFRTYQTLKGVF--PVISTFRPTGYVEIVSYKEN--GH---KVVDTEIY 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
299 VTLMKQVPGAGCPALDAEYSYIS--HNYVGGGSGNGI--MNNKLDNLXGLMTEYNGSTESRR 356
```

```

Db      274 AALTIQNGQOPAVSKSEYSSVHNFLGYSGSGRTSPDSQDNLXLTGKTYSSIB--- 330
Qy      357 YKKGEGDQVIRIRTRNNHLLTSECKQONGYIQTETAYALIGNPNQSPQPLP- 415
Db      331 -RVLDDGQSVASVIERVENKFKLMTKEAKTDONKRIITTEITYNEDLSKSFSEQPNLQOPS 389
Qy      416 KTKETRSADNSYRSSEITTEFPDESQNPFLKVKDKTKQKIIISPTHEWYEPAGEVDN 475
Db      390 RVLTRYDICTNTREETVNIKSDMKNTLL-IETSGIOK-----EYVYYPNGSGNS 442
Qy      476 CPPEYGFTRFVKKIIGTPYDSEFKDPEKFIQYR---SLIGSQ-----SHVTLKIEE 526
Db      443 CPADPLGFSRFLKSVTQKGPDAQSVANKVIHYTQKFPFTGAUYKEVYSKVSSETIDN 502
Qy      527 RHVATQNLNSTLFOYNTDKSELGRLLKQECTGKNGKTSVVKHTYTKQDDTLQOSH 586
Db      503 KIARTFSYVNSP-----TSKSH-GLAKITSVMNQ-----QVYTTREYSESEMETTA 551
Qy      587 SITTHNFTIHSQVRSRYTGRLPSPDPTDKDVTOMSYDKRLTLRLTNGSPYANTLT 646
Db      552 TVTGFDAHMSKAVTSIYTHRQARKVDVNHVITDQSDLLGRITGQIIDGTAREIKRN 611
Qy      647 YDYELNNLQODNRPFPVITTTVDVNGNQLRNEFDAGRHVSQCLKDSG-----DKKF 698
Db      612 YVYQPGDENDFWP-VWIEVDQGVARKTHYDGMGRICSIIEODDQAMGTSIGIYQGY 670
Qy      699 YTHIQQYDEGGRHTSTYSPLYNKGQQDP-----DKHLSKSKSYDMNGQALNTHMS 753
Db      671 RKVLARQYDVLGQLSKEISNDMLWN--LSANPLVRLATPLVTTKYKKGNGNLYSTEYS 728
Qy      754 YGSEKTIQVDPITLTATK-----OLGNSNNVOGKEVTTYTSQOPIQTLFDEAGHL 807
Db      729 DGRLELEHNDITTTITIQGVKGLMNLNQNNP-----EGQASIKAVYPGDTI 776
Qy      808 QSCHTLFRDGMDRVAKKETDAIGQCTIYQYDNYNRVLIQTLTDPGTIVNRKVAPESTDLIT 867
Db      777 YSTRYRVYDGFGRVTEETDAEGHATQIYGDFDRIVKKTLPDGTILLESAYASFHEELIS 836
Qy      868 DIRNGISLGOQTDPGLSRLTQSDGQGVMAVYTSAGNDQCPSTVITPDQGFHYQYQPE 927
Db      837 ALANNQTQGLALVYDGLRIVISDTVGGRKTEYLGPQGDK-PIQSTPSSHKKQNNMDVLYY 895
Qy      928 LDDAVLQVANSNEITQOQFSYNPVTGALLKAVAGOSLTPFI-YYPGRLKEMEN-INDMKMS 985
Db      896 LGSVMSKTTGTDOQNFHYHSGKTGLSLA-SEGSQNTYKSPFSGVLORESEFLANDNPEIS 954
Qy      986 ---YLMTLRGLNGEYTDLTGTIQKISRDTGRVTOIKDSSIKTTLNVDLNRHIGSQYTD 1042
Db      955 SGEYLVTWSGLIQRHKDSFGHNVYSYDAQRLVXTEQDAQYAFPEYDNGRLITTTTKD 1014
Qy      1043 LATGHMLTTYEPFGDNLREIGRLCDSSGHTLIDIOQSLKTOQLANRIVXANGLVQREQ 1102
Db      1015 TTLSQVLTWKIEYDAFDEIRKSLISDPSIOV-ITLSTYKNNQISQRTLSIDGVMMKNER 1073
Qy      1103 YSYSRNKLNYQKCGACGECPTDKYGSIVTONFTYDIAGNITACHTTADGETEDHATPEF 1162
Db      1074 YQYDNNQSLQYQCEGEGSPIDHTGRVLANQOIYHDMQENIKRLDNTYRDKET-VDYHF 1132
Qy      1163 ANPTDPCQLTEVHHTHPMDNIRILKYDKAGRVINIT-DNHGNTENFTYDTLGRL---QN 1218
Db      1133 SQ-ADPQLIRI---TSDKQOIELSYDANG---NLTRDEKQGT--LIYDQNNRLVQYKD 1182
Qy      1219 GQGSV---YGYDPLNRLVSO--KTDTLQCELYRRETMVNEVRNEMRLRLRTGETITIAQ 1273
Db      1183 RLGNLVCSYQYDALNKLTAQVLANGTANRQ-HYASGKTYNTQLGDEALTWSSDQORGH 1241
Qy      1274 QRA-----SKVLLTGTDSQOSVLTGDKONLSQEAYSAGKHKSTANDASILGNGERADP 1329
Db      1242 QSAKKGQSVYYIOYGIDHNSVTYASQENELMALSTTPGFRSLT---SSLPLNKAQYDP 1298
Qy      1330 VSGVTHLNGVRSYDPTLMRFHTPDSLSPFGAGINPYSYCLGDPINSDPSGHLSWGAW 1389

```

```

Db      1299 VTGMYFLGNGRVFNPLAKRPHSDWSPFGRGGINPYTYCQGDPINRIDLNGHLSAGGI 1358
Qy      1390 TGIQMGJAGLLTIAT--GGMALIA---AAGGIAAIASTSTTALAAGALSVTSDITSIV 1443
Db      1359 LGIIVLAGIITIVISIGAGALISAGLIAAGLALASALAVRTVYIGLAADSIGIA 1418
Qy      1444 SGALIEDAPVASSIILGVSMGMAAGLAESAIKGTKLATHLGAFEDGENALIKST-SE 1502
Db      1419 SALSSEKDPKTSGLIMWISAGLISFGISAI-----TFT-----SSIVKSARSQ 1463
Qy      1503 SSRKMGVTRSLDREIVRNEEQYIKDHSR 1532
Db      1464 SOAVSAGVIGSVLEF-----GEVABRSSR 1488

RESULT 2
US-09-492-709A-340
; Sequence 340, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zykend, Judith
; APPLICANT: Olesen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Foreych, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492, 709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: PatsSeq for Windows Version 3.0
; SEQ ID NO 340
; LENGTH: 1426
; TYPE: PRT
; ORGANISM: E. Coli
US-09-492-709A-340

Query Match      3.8%; Score 339; DB 2; Length 1426;
Best Local Similarity 20.2%; Pred. No. 9.6e-17;
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;

Qy      572 KFTYTKDDTLQOSHSTIDNFTIHSQVRSRYTGRLPSPDPTDKDVTOMSYDKLGRIL 631
Db      317 RYITTEAGELL---AVYDRSNTQVRAFTYDAQRGMVAHRYAGRPDMKRYIDTGKRV 372
Qy      632 TRTINSCTPYANLTYDYELNNLQODNRPFPVITTTD----- 668
Db      373 EQNLNAPG-----LSYRV-----LYEQDR-----IIVTSLNRREVLAHTEGAGLKRIVYKE 418
Qy      669 -VNGNQLRNEFDGAGRHSVQCLKDSGDGKFYTHHQYD-----EGGRHHTSYSDYL 721
Db      419 LADSVSVRSGYDAGRLTAQ--TDAAGRRTYEGVLANVSGDITDITTPDGRGRTKEYYND-- 474
Qy      722 TNGRQGT---DPRKVHLSMSKSYDMNGQALNTHMSYGVSEKITVDPITLTATQLOQNSN 778
Db      475 --GNQLTAVVSPD--GLESREYDEPGRLV-----SE 502
Qy      779 NVQGEKVT--TYTPSQOPIQTLFDEAGHLQSCHTLFRDGMDRVAKETDAIGQCTIYQYD 837
Db      503 TSRSGEYRKYDDAHSELPAITTTDAG---STRQMTMSRGQOLAFDGSQGYQTRYEYD 559
Qy      838 NYNREVIQTLTDPGTIVNRKVAPESTDLITDIRVNGISLGOQTPDGLSRLTQSDQG-GRV 896
Db      560 RFGQM-----TAVHRE-----EGISL-YRBYDNRGLTSVYDAQGRE 595
Qy      897 WATYTSAGNDQCPSTVITTPSQFIHYQY----- 925
Db      596 TRYENNAAGDL--TAVITPDGNRSETQYDAMGKAVSTTQGGLTRSMEYDAAGRVISLTNE 653

```

```

QY 926 -----PELDPAVLQVANSNEITQOFSYNPVYTGALLKAVAGSQSLPIYPSGRKME 976
DB 654 NSGHSVSYALDRLVQGGPDGRTQRYHD-LTGKLTQSEDEGLVLMYDESRTTHR 712
QY 977 NINDMKMSYMLTGLGLENGYDTLGTIQKISRDTHGRVTQIKDS-----IKTLLYDDL 1032
DB 713 TVNGEPAEQOY-----DGHGMLTIDISLSEGHVAVHYGDDK 751
QY 1033 NRHIG--SOYTDLATGML----- 1049
DB 752 GRLTGCQYVENPEPTGLMOMHTKAYNEGLANRVTPDLSPEVWLYTSGYLAQMKL 811
QY 1050 ---TTTVEF--DGLNREIGRKLCDSSGH--TLIDQOSWLTQQLANR----- 1089
DB 812 GGTPLVEYTRDLHRETVRSPGMSNAAYELSTTPAGQLOSLNGLVYDRDYGWS 871
QY 1090 ---YKLVNGLVLTQTEQSYDSRNRLNQYKCDGAC-----PTDKYHSI----- 1130
DB 872 DNGDLVVISGPRQ-TREYGSATGRLESVRLAPDLDIRIPYATDPAGNRLPDELPDS 930
QY 1131 -----VTQN-----FTYDIYGNITACHTPADG---TEDHATFKFANPTDPCQULEV 1174
DB 931 TLTVPDNRILAEPAHYVRHDEYGRLEKTDRIPAVIRTDDEBT----- 975
QY 1175 HHTHPDMPDNRILKYDKAGRVINITD--NHGN--TENFTYDTLGR----- 1215
DB 976 HHTH-----YDSOHLVFTYTRIQHGEPLVESRYLYDPLGRBAKXVWRERDLTG 1025
QY 1216 ---LONGGSVYGYDPLNRLVSOQTDPLCELYR----- 1247
DB 1026 WMSLSRKEVYTWGMDG-DRLTWTQDTTRIQYVEGSGFTPLIRVETENGEREKAQRS 1084
QY 1248 --ETMLVNEVRNG-----EMIRLL-RTGETTIA-----QQRASKY 1279
DB 1085 LAETLQOEGSENGHGVPAELVRLDLREBEIRADRVSSBRAMLAQGLVEQLARQV 1144
QY 1280 LLTGTSQOS-----VILTSQKONLSQEA--YSAYGKHSTANDASILGYN----- 1323
DB 1145 EPEYTPARKAHLHCHRGRLPALISBDGTAMSAEIDENQOLNEENPHV--YQYRL 1202
QY 1324 -GERADPVSGVTHLNGYRSYDPTLMBFHTPDSLSPPG-AGGINPVSYCLADPINSRPS 1381
DB 1203 PGQOHDEESGLY--NHRYYDDPLQGRYITOD--PGLKGGWMLYQYPL-NPLQQLDPM 1256
QY 1382 GHSWQWMTGIGMIGIACLLTITATGMAIAGGIAAIAISTTALAFGLSVTSITS 1441
DB 1257 GLI--QTMDDARSG-----ACTGV--CGVLSRIIGSKPDSYADALD----- 1296
QY 1442 IVSGALEDASPKASSILGWYMGWGAAGLAESAIKGSKLATHLGAFADEGENALLKSTS 1501
DB 1297 ---ALKEQONRS-----LCNDMEYSIGYCKDTNG-----KYFASKAETDNLK--K 1336
QY 1502 ESSRIKMGVTRSLDREIVRNEEGOVIKDHSRGYTDNFMGKEQAILVHGDK--DGLPYHT 1559
DB 1337 ESYPLRKKCPTGDRVAAHYTHG---ADSHGDYVDEFFSSDKVLVSKONNLEAFYLAT 1593
QY 1560 EGKMH--NGKRP 1570
DB 1394 PDGRFEALNNKGEY 1407

```

```

RESULT 3
US-09-711-164-467
; Sequence 467, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Onksen, Karl
; APPLICANT: Zwickel, Judith
; TITLE OR INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711.164
; CURRENT FILING DATE: 2000-11-09

```

```

; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 467
; LENGTH: 1377
; TYPE: PRS
; ORGANISM: Escherichia coli
US-09-711-164-467

Query Match      3.7%; Score 331.5; DB 2; Length 1377;
Best Local Similarity 20.8%; Pred. No. 3.5e-16;
Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;

QY 211 LFPQGRGCV-RTBLRPLNQLNSIHNPSLGNENPLTWSFGYTGKGLIGQMTSMTAP 269
DB 126 LFPGB-DGYSRSESLWLR--GGVAKLDEGRALALMQ----- 160
QY 270 GGLKETVNVY-----NNNQ-----HFPPOSANL--PVLPLYTLMKQVPGAGQ- 310
DB 161 -ALPEELRLSPHRYLATNSPQSPWMLGMCERVPDEADVLPALPFRVLTGLVDRFGRT 219
QY 311 PAIOAEVSYSHYVGGSGNGIWNKLDNLGYLMTENYNGSTESRRYKKEGHDQIVRIE 370
DB 220 QTFHREAAAGFSGEIGTVDGAWRH-----FRLVL-----TQGAOR----- 255
QY 371 RTYNNHILTSCKQONGYQTTETAYALIGHNPSQSPQFQPKT--KTETWSADN 427
DB 256 -----AEEARQ-----ALSG--GTEPSAF--PDLTPYTEYGR--DN 287
QY 428 SYSEITETTPDSGNPLTKVIKDKTKQKILISPTHEVYPPAGEVNDCCPBP--YGFT 484
DB 288 GILSLAVMLTHDE-----YP-----ENLPAALVRYKGT 317
QY 485 -----RFYKKIIGPYDSEFK-----DOPEFIQYRSLIGSQSHVTLKIE 525
DB 318 PRGELAVVYDSQKQVRSFYDDKRGMAVHRHTGPE--IRRYD--SDGVTBQLN 372
QY 526 ERHYSATQLNSTLFOYNTDKSELGRLKQTEC--TKGENKTYSYVHKRTTYKQDDTLQ 583
DB 373 PAGLSYTA-----YQEKRIITITSLDRREVLHTQGEAG--KRVYK----- 413
QY 584 QSHSITTHDNFTIHSQSVRGRYGRLEPSDDTDIYQMSYDKLGRLLTTLNSGTPYAN 643
DB 414 -----EHADSVTQSPDA--VGRLEAQDTAAGRTTESPPDVVTLGITIT--TPDGR 462
QY 644 TLTUYELNLIQDNRPFVITTTDVNGQLRNEFPDAGRHVSQCLKDS-----GDGKF 698
DB 463 ASAFYNNHNO-----LTSATGPDGELRREYDELGRLIQETAPDGDITRYRYDNP 514
QY 699 YTHHTQOYDEQHNHTSYSDYLTNGRQCTDPKVLHMSKSYDNMQOLANTHMSYVSE 758
DB 515 SDLPKATEADYGRKMTWRSY--GQLSFTQCSQVTRYDHDREFQMAVHREBSLSQ 571
QY 759 KIVTDP1-TLATKQLOSNNNVQTKGKAVTTPSPQOPIQITLFEAGHQ-----SC 810
DB 572 YRAVDSRGQLIAYKDOGHETRYE-----YNLAGDLTAIYAPGSR 612
QY 811 HTLRDGMVRVREKETAIGQCTIYOYDNYNVIQITLPGDITVNRKXAPSTDTLITDIR 870
DB 613 NGTOYDAMGAVRVTQO--GLTRSMEDYDAGRVIRLTSENS-----HTTFRYDVLDRLIQ 666
QY 871 VNGISLQOYF--DGLSLTQSDGGRVMAVYTSAGNDQCPSTVITPDGFIHYQYPEL 928
DB 667 ETGPDGRTQRYHNDLNGKLRSDEGLV-----TMMHYD-EA 702
QY 929 DDVAVLOVANSNEITQOFSYNPVYTGAL--LKAVAGSGLPIY--PSGR--KXENINDM 981
DB 703 DRLTHRTVGETAERQYDE-RGMLDIDISHISGHVA--VHYVDEKRLTGRGRQYVNH 760
QY 982 KMSYLM-----TLRGLNGYDTLGTIQKISRDTHGRVTQIKDSIKTLLYDDL 1033
DB 761 QTEALLMOHETBRAVNAAGLAN--RCIPDLSLPAVEMWLTG-----SGYLAQMKLGD-- 809

```

QY 1034 RHIGSVTLATGMLTTTVEF--DGLNREIGRLCDSSGHTLIDQOSWLKTOQLANRIV 1091  
 DB 810 -----TPVEYTRRLARETLR-----SGRYELTAYIPAGOLQSG-- 846  
 QY 1092 KLNGVLQRTROYSDYSNRRLNQYKCDAGEPTDKYGHSIYQNTFYDIGNITACHTTFA 1151  
 DB 847 HLNLSLSDYTWNDNGELIR-----ISSPRQ-----TRSYSTGSLTGVHTTAA 893  
 QY 1152 DGTEDHATKFRANTDPC--QITEVHHTHPD-----MPDN-----IRLKYDKAGVIN 1197  
 DB 894 -----NLDIRLPVATDPAGNRLPD--PELHPDSTLSMWPDMRIADAHYLYRDRHGLTE 947  
 QY 1198 ITD-----NHNENFTYDTLGRLONGQ-----GSYGYDPLNRLVSOKTDT 1239  
 DB 948 KTDILPBGVIRTDDETRHRYHSDQHRLVHTRTQYEBPLVESYLYDPLGRPAKAVMR 1007  
 QY 1240 LDCEL-----YY-----RETMVNE-----VRNGEMRLLR---TGETI 1270  
 DB 1008 RERDLTGMWSLSRKQYTWGMDGRLTTIGNDRTRIQTIVQPSFTPLIRVETATGELA 1067  
 QY 1271 IAQORA-----SKULLTGTDSQSVIL-----TS 1294  
 DB 1068 KTORRSLADALQOSGDBGSSVPPVLYQMLDRLSEELLADRVSESRRLWASCGLTVE 1127  
 QY 1295 DKON-----LSQEA-----YSAYGKHKSTAND---AS 1318  
 DB 1128 QMOMQMPVYTPPAKIKLHYCDHGLPLALISKGTTEWCMEYBWKMLNENPHOLOQ 1187  
 QY 1319 ILGNGRADPVSQVTHLGNGYSYDPTLMRFTPTDLSPPG--AGGINPYSCLDGPINR 1377  
 DB 1188 LIRLPQOQYDEBSGLYY--NHRYYDDPLOGRYITOD--PIGLKGMNFGYPL-NEVTN 1241  
 QY 1378 SDPSG 1382  
 DB 1242 TDP LG 1246

RESULT 4  
 US-09-252-991A-28424  
 ; Sequence 28424, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252, 991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074, 788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094, 190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 28424  
 ; LENGTH: 998  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-28424

Query Match 3.6%; Score 317; DB 2; Length 998;  
 Best Local Similarity 22.0%; Pred. No. 2,7e-15;  
 Matches 222; Conservative 129; Mismatches 362; Indels 296; Gaps 47;  
 QY 504 EKPIQYRSLIGSOSHVLTKIEHRYSATQLNSTLFGYNTDKSLRGLAKQTECTGEN 563  
 DB 112 ERASYSYNALG-----LIERADGPRTDVODVTLVAYD-----SRGNL 149  
 QY 564 GKTSVVKFPYTKQDDTLQOSHSTTTDNFTIHSQVRSRYTG--RLFSDPTDIDYTO 621  
 DB 150 TQVYNALQOVTRLGDYDERGKGSITDANGVTSSLA-----YTGVDGLASVSTAGSTTR 204  
 QY 622 MSYDGLRLTRTLNSGTPYANTLTYYDELNNLQDNNRPPIVITTTDVNGQLRNEPDGA 681

DB 205 FDYDAVGQITRYTRGDG-----SWLSYEY-----DDARR--LVAIGNLGERLEYVDTK 252  
 QY 682 GRVSGCLNDSDDGKFFYTHHQO--YDEGR-----HNHSTVS-DYLTNGROOTDP 730  
 DB 253 GNFTAPRIKIDAGS-----LVROQOMAYDBELGRLLRAVAGGQTRSPAYDLNDNPVGETNP 308  
 QY 731 DKVHLSMSKSYDWMGOIANTHMSYGVSEKITYDPIITLATKOLQOSNNVQTSKEVTTYT 790  
 DB 309 RQ--FAHSQAFDALDLVQSGDPLGGKTRLAYD-----AQDMLTEKDPRGVTTYR 357  
 QY 791 PSCQPIQITLFDAGHL-----QSCHTLTRDGMDBRVKRETPALIGOCITYQDNYRVI 843  
 DB 358 E-----YDGLGNLRLVSPDSGTTTFEHNAGVAVIRTRARAGVTEYRYDALNRLV 408  
 QY 844 QITLPGETIYNRKAPFSPDTLLTDIRVNGISLGOOTPDGLSRLOSODGRAWAYTYS 903  
 DB 409 ERKSP-----SDPSLDVQYRYDITLAGNOGIGRLGAIEGARDS----- 446  
 QY 904 GNDCESTVITPDGQFIHQYQPELDVAVLQVANSNETQOFSYNPVYTGALLKVAEGOSL 963  
 DB 447 -----LVRYHD--ERGNLVQVNSIRLDQOTLLDRVTRYRDA-----NQL 485  
 QY 964 TPIYPSGRLMENINDMKMSYLMTLRGLNGYTLDTGTIOKI SRDTHGRVTOIK----- 1019  
 DB 486 LEIGYPS-----GLAICY-----PRNAGQVAVTLAVG 514  
 QY 1020 DSSIKTT-----LNYDDLNRHIGSVTLATGMLTTTVEFG--LAREIGRLKDCS 1069  
 DB 515 DKAPSTLVQIAYLPFEPLOK-----LTWNGGILTSREYDODVYLLQKVG----- 560  
 QY 1070 SGHTLIDQOSWLKTOQLANRIVKLVGLQR-----TEQSYDSNRRLNQYK-CDGAE 1121  
 DB 561 -----FWQSDYQ-----HDANGNIQHRBSLWMTGLYQYDPLRLTEBRVQSG-- 604  
 QY 1122 PTDKIGHSIYQNTFYDIGNIT--ACHTTPADGTEDHATKFRANFPDPCQLTEVHNHHPD 1180  
 DB 605 -----RSYAYDAVGNRTORSNPNASGGTASQDYQYA----- 636  
 QY 1181 MPDNIRL-----KYDKAGVINITDNGNTEFNFTYTLGRLON-----GQGSVGYD 1227  
 DB 637 -PDSNRRLTAIGAQAQVWSDAAG--NLTDQDA--ARKLAYDAQRLQSVLSLGOOVAERLYN 691  
 QY 1228 PL-NRLVSQKTDLTDLCELYRETMVNEVNGEMIRLLRTG-----ETIIAQORA 1276  
 DB 692 ALGGRIVKLPRESITTVLYGPDQGLGEAHDGSGRLRAQVYLMILSLPLATIDADYDA 751  
 QY 1277 S-----KULLTGTDSQOSVILTSK--QNLQEAISAYGKHSTANDAS--ILGNGER 1326  
 DB 752 QGKVGNPFTLLYLHGDHLDTPRLATIDASQALWQSDAFRGELSGSSTQVNLRFPGQY 811  
 QY 1327 ADPVSQVTHLGNGYSYDPTLMRFTPTDLSPPG--AGGINPYSCLDGPINRSDPSG-- 1382  
 DB 812 YDAESGLHY--NFRDYDPTGKRYESD--PIGLSGVNTYGVQGAFLNRIDPLGLAA 866  
 QY 1383 ---HLNQAWMTGIGMGI---AGLLTITATGMAIIMAGIIMAAIYSTT 1425  
 DB 867 IEIDIPKSAVDWIRPGRNRLPAGRL-----GGVLLVASISGATPPADSDT 911

RESULT 5  
 US-09-328-352-6604  
 ; Sequence 6604, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328, 352  
 ; PRIOR FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 6604

```

; LENGTH: 1627
; TYPE: PR1
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6604

Query Match      3.5%; Score 313; DB 2; Length 1627;
Best Local Similarity 20.4%; Pred. No. 1,36-14;
Matches 303; Conservative 177; Mismatches 506; Indels 498; Gaps 73;

QY 75 IGGIGIFGLSVYDRKNSLISLSTGENY--KVIETDKTV-----KIQ 115
DB 369 ITPAMGTEFTHVDAQGLIODISIRTYVSNLQMDAIFGARWYPTTKISRKPKYT 428
QY 116 OKQLDNRPEKDKENYRIIHKSGDLEVLGFNNNAFDLKVPKLNPAGHAIY--ID 172
DB 429 SKKDKHKOYLANGLEYIC-----LDGRAIDLPDLCK-----GOSTYDPIEQ 468
QY 173 WNEEATOPRLNRI-----YDDLGDHIDPLNLEVOGLIKITLLPFGQKEGYTELR 224
DB 469 YTTVTVSDQHLHLAYGDEKRYEKYGEDVRLSYIE-----RKGPFKVALR 514
QY 225 F-----LNROLNGLHNSLGNENPLTWSFGYTPIG-----KNGILGQWITSMTAPGG 271
DB 515 YDHSVTDNKTILSDILFKODDNLALHLLQITPQGLVSDIWTIKNGQLDREVLAS----- 568
QY 272 LKETVNSNNNOGHFPGSANLPLPVITLMKQVPGAGOPALQAEYSYTSH-----NVY 325
DB 569 -----YDYOQO-----DLVQATNEFPAASYVOYTHLLITRYTDLT 604
QY 326 GGGSGNGIWNKNLNLVGLMTEYNGSTESRRYKDEGHDOI VRIERTYNNYHLITSECKQ 385
DB 605 HRGMNLKMDGILPTSKAIEEMADNASRSLBEMDK-----NIRKI-----TYLDVSGNS 653
QY 386 QNGYIQTTEYAYYAIIGHNFDSPQOPQLPKTKETWRSADNSYRSITTTDESNGPL 445
DB 654 TEHYIIDGTYRIYVDPNFE---ECFPRDAKNITLHIAKDSK---TSYTYDERGNVL 707
QY 446 TKVIKDKKTKIISPSTHWEYYP---AGEVDNCPPEPY---GFRPVKKIKIOTPYDSE 498
DB 708 ITTODDCAI-----SYFIDEKNOULTGMDAEGWFOYDSSGNLKEI----- 752
QY 499 FKDDPEKFIQYRSLISGOSHVTLKIBRHYSATOLNSTLFOYNTDKS---ELGRLLK 554
DB 753 ---DPLK-----HETAYV---YNMGLVLTSTIDAKGSKSLKYDDQSLIS 792
QY 555 QTBC---TK---GENGKTVSVH---KFTYTKODDTLQO----- 584
DB 793 YTDCSGKETKQYDERGRVISIEVALNQKEYEFTELTLERREPIIKGLPLNAFGQLEKI 852
QY 585 SHSITTHDNTIHSQVRSRYTGLFSDPTKQIVQMSYDKGRLLTR--LNSGTPYA 642
DB 853 KKHADGTEEHF-IHDAE-----GRLAHVDPKONITTYEYDEAGLILSRIDALN----- 899
QY 643 NTLTYDE---LNLQODNRPPFVITTTVDNGNQLNEFDGAGRHYSQCLKDSGDGKF 698
DB 900 HKLKTKMDRIGRLRLRLNE-----NGASVYFPFDVASRLV---KEIDPFGKE 943
QY 699 YTHHQOYDQO-GHHTS---TYSDYLTNGROQTD-PKVHLSKSYDNMNGQIANTWSY 754
DB 944 TVVH---YDEKSGQLATSIENVASVYQODLKDRAPAKRIQOPI---FDSMGRLEORTAGY 997
QY 755 G-----VSEKITVDP-I-TLATQQLQSNNSNNVOTGKEVNTYTSOQPIQITLLEDA 804
DB 998 GHYGLBLEEKQTEFPAYDMGRILIIQANNAOSNLO-----WFDAA 1037
QY 805 GHLSQCHTLTRDQWDRVKEITDAIGQCTI--YOYDNRNRYIQLTLPDGTIVNRKAPFST 862
DB 1038 GNLVQEH-----QQDYKINKTAVMKQYDEIDNDRIKTRPRGQYID----- 1078
QY 863 DTLITDIRVNGISLGQOTFPGLSRLTOSQDGRVWATYSAANDQCSYITTPDGFITX 922
DB 1079 -----W-LTYSGHVQ---SLIVNGODFVSP 1100

```

```

QY 923 QYQPELDVAVLQVANSNEITQOFSYNPVTGALLKAVAGOSLPTIYPSGRLKMENTINDMK 982
DB 1101 -----ERDDLREJA-----RHYANGVSQEOQYDLAGRLKQOM----- 1134
QY 983 KMSYLTWLRGLNGYTD-----LTGTIOKISR-----DTHGRVYQIKDSI-KTTLNY 1029
DB 1135 -----LSHENGYQNVQRHNNALQOTQLVQYQVKTGBELAIRDTRRGNIAYKY 1187
QY 1030 DDLANRHISQVYDLAAGHMLTTVEPDGLNREIGRLKDCSSGHTLIDIQSWLTKQOLA-- 1087
DB 1188 DPGRLLE-----ASSKKEKTFSPDPASNIL-----DSYHSQKVQSHQKLDSETS YG 1235
QY 1088 -NRIVNLGVQOR--TEQYSYDSRNRLNOYKCGAGCEPTQKXHSIYTONFTYDIYGNIT 1144
DB 1236 YNRIV--NNVAKSYLDQOYQYDAVQGLIRKTSQGLD-----NLEMDVIGRMV 1281
QY 1145 ACHTTPADGTEDHATFKNFANPTPCOLTEVHHTPPMPDNRILKYDKAGRVINITDNGN 1204
DB 1282 KSRNS-----QYTAEY-----RYDALGRRIQKMSKHH 1309
QY 1205 TENFTYDILGRLONGGSVGYDPLNRLVSQKTDITLDCELYRETMVNEVRNGEM--IR 1262
DB 1310 T-----GOEQN---IIVGMDG-DTLAYESTBELTKHYIYERKDSFVPMLOAVYLSPIR 1357
QY 1263 LKRTGETI-----IAQORASKVLLTGPDSSQSVILTSDKONLSOE-----AYS 1305
DB 1358 LHQTPMDSRPPVNIHDPMLKTEKGEKEDDWFYHCDHGTQEMTDHTGAILIKAEYK 1417
QY 1306 AYGRKHSNDASII-----LQNGERADPVSGVTHLNGYSYDPTLNRFPHTPDS 1355
DB 1418 AMGECKABEAKSNFENSEIISNNIRPQGOYFDEBEGLAH--NRVRYSPYVGRFVSKD- 1474
QY 1356 LSPFG-AGGINPYSYCLADPINSRSPSGHLSQAMTGICMGING 1398
DB 1475 --PIGLGGNNVYVA-KNPTIWDSKGLCSTIANRLG-GVNG 1514

RESULT 6
US-09-543-681A-5329
; Sequence 5329, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5329
; LENGTH: 1586
; TYPE: PR1
; ORGANISM: Proteus mirabilis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (1576)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-543-681A-5329

Query Match      3.4%; Score 306; DB 2; Length 1586;
Best Local Similarity 22.1%; Pred. No. 4,4e-14;
Matches 243; Conservative 129; Mismatches 385; Indels 342; Gaps 59;

QY 499 FKDDPEK-----IOYRYS-----LISQSHVTLKIBRHYSATOLINS- 537
DB 486 FEANPHKINTLRMKSGADRNHNLHRYANDSLVQIHDAVYLTDIRHVDLTTQRLQSV 545
QY 538 -----TLFOYNTDKSELGLLKQTECTGKNGKTVSVHKTYYTKODDTL---QOS 585
DB 546 TRHGOBEKTLVYTYTDAQO--RLVQVTVNADK-----RVTRRFQWDESGIMAMHQYA 596

```

QY 586 HSIITTH-----DNFTIHSQVRSRYTGRLLFSPTDITKDIYVMSYDKLGRLLTFTLNSGT 639  
Db 597 TGVSSHTRKQRFDAFTIEDNEPREVVEHMKDKRCLEHTELTYDLAQTITLTVENGGE 656  
QY 640 PYANTLLTYDVELNLLQDNRPPEVITTTDVANGNQLNEPFGAGHVSQCLKDS---DGDG 696  
Db 657 -----TFRRMNEQQOIIIEYTNALNETMFEWD-----TSRLAKKAIPAGSSE 699  
QY 697 KFTYITHQOYDEOGRHHTSTYSDDLTVNGROQTDDKXHLMSKSYDMWQO-----IANT- 750  
Db 700 WGYT-----YDERG-----NLQMTDPEQO-----OSTCYD-WDKQFAPIYQTL 737  
QY 751 -----HMSYGVSEKI--TVDP--ITLLATKOLQSNMNVQSGEYTYTPSQOPIQT 799  
Db 738 PNGAMMEVNEHEDIRVIDLPLHITRLAMDQGLGQVDAKGNETHRYNARGQLIE 797  
QY 800 LFDAGHLQSCHTLTRGDMRVKRETDALIGCTTYQDNINRYQITLPDGTIVNKKYAP 859  
Db 798 QRDSCGYPT--TLTYDWMQGLRSITNQNEMTYTSEAGLLTECLPDET--ENRYDY 852  
QY 860 FSTDTL--ITDIRVNGISL-----GQ-----OTFDSLRLTOSQDGRV 896  
Db 853 DATQQLVIGTIDAGRHILLRRNRGQVYARDPAGHMLHFHYDTFGKQALENQ--GEQ 910  
QY 897 WATYSAGNDQCPSTVITPP-----GQFIHQYQO-----BELDAVLQVANSNE 939  
Db 911 YRFPYDLNHR-----LTDEHDLGGQKHGYDVMGNVQTKTPGPSIDTPI--PLSPQ 962  
QY 940 ITQOFSINPVGALLKAVAGQSILPIYPSGRLKMERINMKKMSYIM-----TLGLE 994  
Db 963 VT--TFGYDKV-----GRLLFRENA DY--RTEYLQPLSVTLRRVP 999  
QY 995 NGY---TDLTGKIKISND-----THGRVQO--IKOSIKITL--NYDDLANR----- 1034  
Db 1000 MAVMEHEKRTITKARVEYQDALTFTYDKVQGLVREBASRGYQHNYDVLGKITTELPHQ 1059  
QY 1035 -----HIGS---QVTDLATGMLTTVEF--DGLNREIGRKLCDSSGHTLIDQSWLKT 1083  
Db 1060 RAFELYVYGSHLQOTQWRDNEQLVLAEYGRDLRHRETLR-----TSG--ALDNEGYDCR 1114  
QY 1084 QQLANRYK-----LNGVLQRTQOYSYDSRNRL-----NOYK 1115  
Db 1115 GRITHQVARKQMNASQFVTPVIDR--RYRMKRNQILERSVSQGTGEVFTAGHWYHSYQ 1172  
QY 1116 CDGAECPFDKXGHSIVTONFTYDITGNITACHTTFADGTEDHATFKFANPTDPCQLTEVH 1175  
Db 1173 YDPLQGLTALHIG--SVQTEHFLYDAANLL-----TRPH-----TRAP 1208  
QY 1176 HTHBDMPDNIRLKYDKAGRVINITDN-----HNTEN-----FT 1209  
Db 1209 HNQVQSGDKFPRYDGFGRMVSRYEKSSSGQRHYHDSHRIIAVDIDQGLGYQABAYR 1268  
QY 1210 YDTIGRL-----QNGQGSYVGDPLN--RLVSGKTTLLCELYRRTM 1250  
Db 1269 YDILGRIRLEKRLWKASAIANTVTVYQHBPDEYTFGWMGMRVSHSASAHPHTTVH--- 1335  
QY 1251 LVNEVRANGEMIRLRTGETTIIAQORASKVLLGTDSQOSVILTSQKMLSQEASVAYG-- 1308  
Db 1336 AYNDSYTPPLARIECT--DNPLNPOALYYTHSSLSGLPEALNBSGEIYVWQGSAMWHL 1384  
QY 1309 -----XHKSTANDASILGNGERADPVSGVTHLGNGYSYDPTLMKFHTTDPDLSLPG-AGG 1363  
Db 1385 QORQRTPTSTFNRBQQLRFQGGYFDKETGLHY--NTRFYAAPDLGSGFTQOD--PIGLAG 1439  
QY 1364 INPYSYCIAGDPINRSDPSG 1382  
Db 1440 INLYAYVA--PNPLTWDPWG 1457

RESULT 7  
US-09-543-681A-7560  
; Sequence 7560, Application US/09543681A  
; Patent No. 6605709

; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OR INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709,1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 7560  
; LENGTH: 1439  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
; US-09-543-681A-7560

Query Match 3.3%; Score 294; DB 2; Length 1439;  
Best Local Similarity 20.3%; Pred. No. 3.2e-13;  
Matches 251; Conservative 141; Mismatches 386; Indels 460; Gaps 62;

QY 347 YNYGTESTERYK-----DKEGHDQIVRIERTYNNYH-----LTSECKQONGYIQ 391  
Db 376 YHPPFDKYSRLSHVDVNGEQRH-----YNEHQQLHNGCGDLNTECEYOS--FQ 429  
QY 392 TTE-----TAYYALIGHNFDQSPQF-----QLPXTKET-----WR 423  
Db 430 LAETVSRLLTAVYQVNPHQIRRLCAVFNESAQLRVEQOTNHPYRQFGMTDAGVAMH 489  
QY 424 SADNSYSELTETTPDSSGNPLTKVJKDKTKQKIISSTHMEYPPRAGEVDCPEPYGF 483  
Db 490 SDKYGLSEKRMALSEN--LWRVLENKTS-----GESYRLBYDINL 531  
QY 484 TRFVKKIIQTPYDSEFDDKEFLQYRYSILGSSOHTLXIEBHYSATOL----- 535  
Db 532 TR-----TAY--WHDGSTFWQLNH-----DHQIHVDRGTGKTLIMDEFGLPC 575  
QY 536 ---NSTLFOYNTDKSELGRLLKOTECTKGNGKTVSVHKEFTYTKODDTLQOHSITTHD 592  
Db 576 GCRNALGHTHISEWDALGRLLSITD--GNGNQF--RWOQONERERL-----IT-- 619  
QY 593 NFTIHSQVRSRYTGRLLFSPTDITKDIYVMSYDKLGRLLRFLNSGTPYANTLLTYDELN 652  
Db 620 -----TFWPDNTE--SRLAYDSLGLIKEL--SPHQTIRYRDFK 656  
QY 653 NLADNRPPEVITTTDVANGNQLNEPFGAGHVSQCLKDSGDKFYTITHQYDEQGRH 712  
Db 657 TTL--RP-----TIRIDAQGRSEFLMNR--COLLRHDDCSGR--QHIMCWDDSGRV 703  
QY 713 HTSTYSDDLTVNGROQTPDKVHLMSKSYDMWQOIANTHMSYGVSEKITVDPITLTATQ 772  
Db 704 VSQ-----TNAIQE-----ATEYQYDEVGHILARI-----ILPD--NSTVQ 736  
QY 773 LOSNSNVQSGEYTYTPSQOPIQITLFDAGHLQSCHTLTRGDMRVKRETDALIGCT 832  
Db 737 LAMNAAGLLTHQNDNTPQCMQY-----NAFGVTEIDKLAHRI 777  
QY 833 IYQDYNINRYQITLPDQ--TIYNRKYAPSTDTLITDIRVNGISLQOQTFDGLSRLTQS 890  
Db 778 HHYNMAGALISINMAGRLNLRD-----AEDRLVEIRPDE--TLQYTVNAVAGRLVEE 832  
QY 891 QD--GGRYVA-----YTSAGNDQCPSTVITPPGQFIHQYQBELDAVLQVANSNETQ 942  
Db 833 AHLCDRVFTSNPRTILLDYDAAGMLVGRRETLTD-----RQYQWDSNMRL--VASKOPNQ 886  
QY 943 -----QPSYNPVTGALKAVAGQSILPIYPS 970  
Db 887 RGLEMGLOANQVHTFYDALGRITREQTGDDIVERNVDLNLRLTLPQGSILMKLYGS 946  
QY 971 GRLKMNINIMKMKSYLMTLRLGLENGYTLGTIOKISRDTGHGVTQIKDSIKTTLNVD 1030  
Db 947 GHATRAIN-----HLVDSR-----SQLITEFRD 969  
QY 1031 DLNRHIGSQVTDLATGMLTTTVEFDGLNREIGRKLCDSSGHTLD--IQOSWLKTQOLA 1087

```

Db 970 DLHREIS-----RTQBELTQYRQYDKLGRITSTPSSNDKQHPNLNGITLWAKMWYDQ-- 1021
Qy 1088 NRIVKANGVLQRTQYSDSNRLN-----OYKCDGAECPY 1123
Db 1022 GNLCAMEDTYRGWWEYLYDSRQLKVAASENLAMLYRADNMLERPOSEMNAHSPT 1081
Qy 1124 -----DKYGHSLVTQNFYDIGNITACHTTFADGDEHATKPAFNPDPQLTEVHH 1176
Db 1082 LELSPQDKL-RQPGWHYQDAYGVNVA-----RRYRN-----OSSQTY- 1120
Qy 1177 THPMPDNIRLKYDKAGRVINITDNHNTENFTYDTLGR----- 1215
Db 1121 -----AYGDNRLLV-IANNGIKAQYHDLGRIRHKTKVENRESQVAKQETHF 1168
Qy 1216 -----LQNGGSVY-----GYDPLNRLVSKQDITDLCELYRETMVLN- 1253
Db 1169 IMQGLRLLOQODINTGKQTCYEHSYTPPLAVIVKQSSGF-----HYWHHCIDINAPL 1224
Qy 1254 EVRNGEMIRLLRGTETIIAQ-----RASKVLLJTGDSQSVILNSDKONLSQEAISA 1306
Db 1225 EVTNAAQ-----GNTIWSGKTERFGFVRSSEPLSPSSPER---VMASFQNLK- 1268
Qy 1307 YGKRKSTANDASILGNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFG-AGGIN 1365
Db 1269 -----VAGQYFDNETGL-HF-NTFRFYDPQIGRIMPD---PIGLGGIN 1308
Qy 1366 PYSTCLADPPINRSPSG-----HUSWQAMTGI 1392
Db 1309 LYQYA-ENPLGWIDPMGLSSQEMVRVRHHTSVSEGLEGI 1345

```

## RESULT 8

```

US-09-252-991A-23805
; Sequence 23805, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,768
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23805
; LENGTH: 1626
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23805

```

Query Match 3.3%; Score 292; DB 2; Length 1626;

Best Local Similarity 21.1%; Pred. No. 5.7e-13; Indels 320; Gaps 46;

Matches 204; Conservative 125; Mismatches 316;

```

Qy 597 HRSQVRSRY-----TGRLFSDTDITQMSYDQAGRLTRTINSCTPY----- 641
Db 633 HTTQVLAHYCNCRGQILBAGNMLYEABRYRVDABNVILRQLAGASPFMEWEGSKQA 692
Qy 642 -----ANTITYP-YELAN-----LQDNRPFPVITTTDVNGNQLRNB 677
Db 693 RAVHMAASFQOMSRVYMNDSVTALNADGSEBVYVHDNAR--LVROYDPGGETLRH 750
Qy 678 PDGAGRHSQCLKSDSDGKFYTIHQDYEOGR-----HHTSYTS----- 718
Db 751 YDKRGQVAA--RDPDG-----TITTEYRVDAGRLMLLAABGEPTYSITFDGFSVRR 803
Qy 719 -----DYLTNGR-----QOTPD--KVHLSMSKSYDNMGQJANTHMSYGVSEKTIIVDPTL 767
Db 804 GEAQWKYERNAQGDITQTDPFEGNVTHY-----AYDHRGCLVEITHADGSLHQLTWNPLGQ 859

```

```

Qy 768 TATQLOSNNSNVOTGKRYVTTYTPSQOPIQITLFDENGLQSCHTLTRDGMVRKETA 827
Db 860 LIEQLPDGS-----VRRYRYDTLGRQITRDESGAI----- 891
Qy 828 IGQCTIQVQNNYNNRVIQITLPDGTIVNRKYAPFSTDTLINDIRVNGISLGQOTPDGLSRL 887
Db 892 -----TRQWMAAGRLSITLPDG--ASRMT-----KRN--AYGKVT----- 924
Qy 888 TQSDGGRVWAAVYYSAGNDQCPSTVITPDGQFIHYQYQPELDVAIVASNEITQAFSYN 947
Db 925 SECDQGRMRYEX-LDDLHLVSRINPDGSQRLYRX--ENARLLSEIENERGERYRLD 981
Qy 948 PYTGALLKAAVEQSLPTIYPSGRKLMENINMKMSYLTMLRG--LE-----NGTDL 1000
Db 982 -----YHNGLLSQETGFGRRRAYRYDLKGLLEKTEYDDDSGL 1022
Qy 1001 TGTIQKISR-----DTGRVYQIQQSSIKTTIANYDDINRHIGSQVTD 1042
Db 1023 RTTYQRDSTGRLLAKTLPDGNRVDRYRDTLGRVAVVDGTWPLAVEYDLDNR----- 1074
Qy 1043 LATGHMLTTVE--FDGLNREIGRKLCD-----SSGHTL--DIQGSMLKTQOLANR 1089
Db 1075 LVREHQMATILHAYVDALGQLIHCRLLPDGNRVDRYRYGTGTLSDILMGSLTRHQFGSG 1134
Qy 1090 IVKLNGVLQRTQYSDYSRRRLNQYKCDGAECPDKTGHISIVTONFTYDIGNITACHTT 1149
Db 1135 RERORQGSLLSYTHDEQGRLLAHQVSORO-----RHLYORQYRYDASGNLAA----- 1183
Qy 1150 FADGTEHATFKFANPTDPCOLTEVHHTHPDMPDNIRLTKDKAGRVN----- 1197
Db 1184 IEDSRKQIRSFHY-DPLD--RLLGVRGETPE-----SFVNDPKANLAAQGGPAPROMEV 1235
Qy 1198 -----IT-----DNHGN-----TENFTYDTLGR-----LQNGGSVYGYD 1227
Db 1236 RGNRLTQGRHFDYDAGHGLVBERRGYQKLVTEYSYDCQHLIGVSLPDGRQVYRYD 1295
Qy 1228 PLARLVSKQDITDLCELYRETMVLNVEVRNGEMIRLLRGTETIIAQ--QPAKVLTLGTD 1285
Db 1296 AFGRIARLKVQDQ-----RNTFELWL--GERLLAESGDHRYRYLYLEPD 1336
Qy 1286 SQOSVILTS-----DKQNLQSE-----AYSAYGK--HKSTANDA 1317
Db 1337 SFRPLALDSEGEQVEPCYQDHLCTPQELTRADRLCWSARIYAVGVKLADIABVD 1396
Qy 1318 SILGNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFAGGINPYSCIGDPINR 1377
Db 1397 NPLRFQGYFDAETGLHY--NRHRYVNPSTGRYLTDPPIRL--AGGLNSYRY--VPNPTGW 1451
Qy 1378 SDPSG 1382
Db 1452 VDPLG 1456

```

## RESULT 9

```

US-09-328-352-5542
; Sequence 5542, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5542
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5542

```

Query Match 3.2%; Score 285.5; DB 2; Length 1596;







```

QY 1084 QQLANRIV-----KLVGLOR-----TEQSYSDSRNRLNOYK-CDGACPTDKY 1126
D 552 YOLLROKRVQWQSYQDANGNIOQHSHLMGTLIDYQYDPLDKLTERGVGG-----604
QY 1127 GHSIVTGNFTYDIYGNIT-ACHTTFADGTEDHATFKCANPTPCQLTEVHHHTPDMENDI 1185
D 605 -----RSYAVDAVGNRTQSDNPASGDTASSQDYQYA-----PDSN 640
QY 1186 RL-----KYDAQGVINITDNGHNTENTYPTTLGRLOQ-----GQSYGVYDPL-NR 1231
D 641 RLTAIGAQAATSDAG--NLTQDRA-ARKLAYAQAQLQSLSLGGQVAEYRNVNAGOR 696
QY 1232 LVSOQTLTDLCELYRETMVNEVRNGEMIRLRTG-----ETIINQDAS-----1277
D 697 IVKLTPSSVTYTYLGPBGQLGEAEHDSGRKLAQYTLMLDSLPLATIDIDYAOQKVG 756
QY 1278 --KVLITGTDSQOSVILITSDK--ONLSQEAAYSAYGKHKSTANDAS--ILGVNGERADPVS 1331
D 757 NPTLLYHGHDLDPRLATDASGQIAWQMSDAFGRBALSGSGSTQVNLRFPGQYDAES 816
QY 1332 GVTHLGNYSYD-----TLMRHTPD-----SLSPGAGGIMP--1366
D 817 GLRY--NYFRDYDEYETGRVYESDVEETLRKLNPEMTFLNSGESMLQATPYMEHGFPMH 874
QY 1367 -VSYCLGDPINRSD-----PSGHLSP-----QAWTGIQMGIAGLLTIAT 1405
D 875 NYTYSNNPFAKSKHGLSPFTNLIYTPDNTCTCLECKKTKTGKSKP-----925
QY 1406 GGNAAIAAGIAAIAIASTTALAFGALSVTSDITSIVSGALBDASPRASS 1456
D 926 -----LVGALCSKATTPF-FGGVVCNSTIWMICGASCSQECNRAPS 965

```

```

RESULT 11
US-09-543-681A-4476
; Sequence 4476, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543, 681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128, 706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4476
; LENGTH: 1665
; TYPE: PRN
; ORGANISM: Proteus mirabilis
US-09-543-681A-4476

```

```

Query Match 3.1%; Score 271.5; DB 2; Length 1665;
Best Local Similarity 20.7%; Pred. No. 2,4e-11;
Matches 233; Conservative 148; Mismatches 350; Indels 395; Gaps 66;

```

```

QY 508 QYRSLIGSQSHVTLKIEBRYSATQULNSTLFOYNTDKSELGLKQT-----EC 558
D 608 QHRIDPIYNESQSLVSV-KRHIDGIALIL-----EYQDDK-LVVICSESTTSPRDFVRC 659
QY 559 TKENGKTVSVVHKFTYTKODDTLQOSHSTTHNFHISQVARSYTGK-----LPSDITD 614
D 660 EYVTHG-----VLSQCHAYQVNH-----LWHRYSPEGLAVAWGDTD 695
QY 615 TKDVIQMSYDKLGRLLTRTLNSG-----TPYANTLY-----DYELNLTQDDNRPP 661
D 696 STELT-IDYDEGCRVVAHSPSGFWMDFIYDYQMTTYIDBGGFSRYVNDN-----750
QY 662 FVITTT-----DVNGQNLNRPFGAGRHVSQCLD---SDGSGKFY 699
D 751 -LVTRITDPLMRETYTEMEQRKIAEINIEIGERTY---GVHYVGLLAIYIYLPDGKAIY 806

```

```

QY 700 TIHQYDEQGR--HHTSTYSDYLITNGRQCTDPDKVLNSKSYDNMGQIANTHWSYVS 757
D 807 -----DYNQDGLHFTLSAFD-----EWQSLYDENGNTL-----836
QY 758 EKITVDPITTLATKOLSNNSNNVQTKREVTT-----YTPSQQPIQITLFEAGHLOS 809
D 837 --IYTPD---QGRQVVEYSGHGBELKAITPNGAQWQYHNPAHLKTT--NRYQSTE 889
QY 810 CHITLRDGMVRKEDTAIQCTIYQ-----YNNRVY 843
D 890 YHS---DELRLIHYTDLALNHTTRYQYSTERASTNGSLSKILLPDGVEQOIIDYSEBRVI 946
QY 844 QITLPDGTIVNRKAPRSTDTLTDIVNGISLGGQFFDGSRLTOSQD--GGRVAAVYX- 901
D 947 AVTDESKTRIRYGP--DYLLAMIRPDSEI--REFYDSILTKKVVYVANGVEYLYERD 1003
QY 902 SAGN-----DOCPSTYIT--PDGFIHYOYQPELDVAVLQVANSN-----939
D 1004 KAGQIIREVDPTGREICRYRDLGRRIATATYVPMHBLRWY---NESGLVEQSEWPFED 1059
QY 940 -----ITQOFSPNPTGALLKAVAGQSILPIYYSGLKMNINDMKMSIYLTPLRG 992
D 1060 EBSRCLSTAOYSYN-ARQOLIKATNPDSVFEEDYDQGRLCSEIRINE-GEIVHQW----1113
QY 993 LENGYDTLGTIOKI-SRDTH-----GRVTOIKDSIKITTLVNDLNRHIGSOVTLAT 1045
D 1114 --NEADNTIALTRFGRRELIHYAFGALGELTSLD-----VNOHAPLOFS-----1154
QY 1046 GHNLTTFVBPGLNREIGRK---LCDSGHT---LDIQSMWLTQOULNRIYKLVGL 1097
D 1155 -----YNAVQELRLRSRAGFVNSHYTATGLAHQRAGGTGFQF-----LQSIQ 1199
QY 1098 QRTBQ-----YSYDSRNLNOYKCDGACPTDKXGHSIVTONFTYDIYGNITACH 1147
D 1200 AHPQOPEPCTDVHRSYQYD-----RAYNVVGIK--DDRWKQT---RYHVNANDQITETQ 1248
QY 1148 TTFADGTEDHATFKCANPTPCQLTEVHHHTPD---MPDNIR-----LKYDKAGRYINI 1198
D 1249 YSPQWKNQDE-KFOYDNNLN---ITF-HLTPSSSWVPSDAQMLQDFQOQAGRYTRR 1303
QY 1199 TDHNGHNTENTYPTTLGRLO-----NGQGSV-----YGYDP 1228
D 1304 YTKAG-YQDHYDVNGRLAKKIVHTRGFRPREWRYLNTQNLJACPTPGDCWHTYYDA 1362
QY 1229 LARLVSOQTLTDLCELYRETMVNEVRNGEMIRLRTGTTI-----AQO- 1274
D 1363 FGRRLS-KTQTVSDLAHIDPLPQIKITTRRYLMSGDOLIEPPIYADGTLANAOV 1421
QY 1275 ---RASKVLLTGDSQOSV--ILTSDKONLSQEAYSAYG-----KH 1310
D 1422 QMLYQPELTPYIARYOQGLHYVYTDHQGPREFISGQASWAGRLNTWGMQWFRYRD 1481
QY 1311 KSTANDASI-----LGVNGERADPVSQVTHLNGYRSYDPTLMEHPTDLSLSPG-AGGIN 1365
D 1482 GKXENDPNYTECPFRFAGYVEDESGLY--NFRFRYDRETGQYLSDD---PGLGLGGLN 1536
QY 1366 PYST--CLGDPINRSDPSGHLSWQAWTGIQMGIAG---LLTITAT 1405
D 1537 PYGVVHC---PTGMWVP-----FGLAGCCDDELIVLAT 1566

```

```

RESULT 12
US-08-851-567B-61
; Sequence 61, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Patig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.

```



FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 5545  
LENGTH: 804  
TYPE: PRN  
ORGANISM: *Acinetobacter baumannii*  
US-09-328-352-5545

Query Match 2.7%; Score 235.5; DB 2; Length 804;  
Best Local Similarity 20.8%; Pred. No. 4,4e-09;  
Matches 147; Conservative 92; Mismatches 253; Indels 215; Gaps 30;

QY YTHHQYDEQGRHHTSTEDYLTNGQQDPDPRVHLSMKSXYNMQIANTHS----- 753  
DB FTLLQTEYDAQAGLRSTQAGRYVKKYET-----VSFSTCL--GNLASFQMSKVC 71  
QY 754 ---YGVSEKTVDPITLTKQLOSNMNYQTEKVTTPSOQ--PIQITLPEAGHLQ 808  
DB LSNVVLGDQVGLPIEV-----DVGNGKEITQFAGYKRGITQVYKLANGA----- 116  
QY 809 SCHTLTRDGMVRAKETDAIGQCTIYQDYNRVITLTPDGTIVNRKVAPEFTDITLD 868  
DB 117 -TETNIVDDEGNITQHTDADGVISRQYDAGRLL--YIDPIVGLNYSFTYDGLTVS 171  
QY 869 IRNG-----IS--LGQTFP-----GLSRLTQS- 890  
DB 172 RVTGGGQLRIEKNYDGLLISEDKISNKSIIINSKRYAFGNLIPKSNPGSAITSGT 231  
QY 891 -----QDGRVNAVYTSAGNDQCPSTVITPD-----GQFIH----- 921  
DB 232 TSSVDVDRPIYNDNSVTTYCQSCGKGTGAVQOTDPSGTESNLAAGDPSADLKT 291  
QY 922 -----YQYQPELDAVLQ--VASNEITQPSY-----NPVTGALLKA 956  
DB 292 LVARKGTGDSVFQTTTEFENALLKPKVAVSGSTQSYNSNTLATEKONSISGKTFK 351  
QY 957 VARGQSITPIYVSGRKMENIMDKMSYLTMLRGLNGEYGT--DLTITQIKISR----- 1009  
DB 352 YDGTGRTSITHPDSVETIKYFQKDLIASRTWREVEITYSYSLAGRLKTTNNANISEA 411  
QY 1010 ---DTHGRV-----TQIKDSIKITLVNVDLNRHIGSQVTDL-----A 1044  
DB 412 FDDITVGRVISHQKINANDTNNSVYVRYGNQNL-----QVTSIQYPNKGSVNLSONA 466  
QY 1045 TGHMLTTVEPFDGLNREIGRKLCDSSGHTLIDQSMWLTQOLANRIVKLANGVLQR--TE 1101  
DB 467 LGEVTSIPNVISLNVYARQQLTVQANTDMLSYTNDGSLNNISATS--LEKCVLANV 524  
QY 1102 QYGVDSNRNLNOKYKCDGAECPDKYGHISYTONFTYIYGNITACHTTPADGTE--DHATF 1160  
DB 525 DYGVYDKLNRNKKL-----SDKGS--VYNATIDRYG-----TGLMSTVELDQARY 567  
QY 1161 KFNAPTPDPCQTEVYHHTHPMDPNIRLK-----YKAGRVINITDNHNGTENFTY 1210  
DB 568 QYGVYND--DITKY-----NITSKSTVAPAITTYVANTSRILASVSGSTISFTY 616  
QY 1211 DTLGRLONGGSVYGYDPLNRLVSQKTDLDCBLYYRETMVNEVRN 1257  
DB 617 DAMGNVTNDGVRTLTLYDNYSRLLNKGNETY--LYNADGLRVAVRND 660

RESULT 14  
US-09-252-991A-31032  
Sequence 31032, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 31032  
LENGTH: 1128  
TYPE: PRN  
ORGANISM: *Pseudomonas aeruginosa*  
US-09-252-991A-31032

Query Match 2.5%; Score 223; DB 2; Length 1128;  
Best Local Similarity 19.6%; Pred. No. 7.5e-08;  
Matches 258; Conservative 151; Mismatches 467; Indels 442; Gaps 68;

QY 235 NFSIGENPLTWSEFGYTP-----IGKNGILGQMTSMTPAGLKEVYVNSNNQGH--FP 288  
DB 63 DFLAAPDPTFPARGYLLSNPRIGRLG-RGMWLPGBSMHLESDACVLVDAQGRRIQFP 121  
QY 289 QSANPLVPLVYTLMKQYPCAGCPALQAEYSYTSHTYVGGSSNGIWNKKLDNLYGLMTEYN 348  
DB 122 ALA-----PGA-----QHYS-----GSESLW----- 137  
QY 349 YGSTRRYKDKGHDQIVARTYNNYHLLTSECKQONGYIQTETAYYAIIGHNPSQ 408  
DB 138 -----LRGGSSGGBADAMR--GRMAVPAE-----LQIQESVVLVSGHST-- 177  
QY 409 PSQFQPKTETETWRSADNSYRSEIT-----ETPDESQNPYTKVYKDKK-----TQ 455  
DB 178 -LHFQ-----RCPDGIMRLQASPRGAGYRTEFRWSGRGLLTGRDSAGSYALVYQ 227  
QY 456 KITSPTMHTYTPAGEVNDQCPPEYGFTRFVKKIIQTPIYDSFKD--DPEK-----FIQ 508  
DB 228 QACEPS-----EGDD-----GLRLP--GVILASHDGPPIYDQSGELMVLVR 269  
QY 509 YRSLSGSGSHVTLKIEBRYSATOLL-----NSTLFOYNTD--KSELGRLLK 554  
DB 270 YQPSDSGD--LIAYRDLQGVYVFAHREHMLVAHGEGLLEVRYEMVHAHGHVAK 325  
QY 555 QTECTGKNGKTYSVVHKPYTKDDTLQOSSHITTHDNFTIHRSGVRSYTRGLPSDIT 614  
DB 326 QIAG-----GLRTPRYLR--DATEVSDSLGVEREFAGEGQRWTLVLRADGS 375  
QY 615 TKDIYQMSYDKGLRLTLTSLNSPTAYNTLYDYELNLIQDNNRPPIVTTTIDVANGQL 674  
DB 376 R-----SEFYDILFGRLL-----VAMRDPILGREY 398  
QY 675 RNEFDAGRHVSGCLDSQDQKFTYIHHQYDEQGRHHTSYSDVLTNGRQOTDPPK-- 732  
DB 399 RRRRDQGRWLE--ESPGRARY-----RKRVDEBTGL 429  
QY 733 -VHLSMSKSYDMNGOIANTHMSYGVSEKITVDEITLTKQLOSNMNYQ--TGKERTTY 769  
DB 430 LVLE-----DAMQRRWTFERDER-----GNATTVRGAPGSTRYAY 465  
QY 790 TPEQOPQIQT-LFDEAGHLQSCHTLTRDGMNR--VRKETAIGQCTIYQDYNRVYQI 845  
DB 466 EDPRLPDRPRIYDPGGER-----RLEWNRGLAALLTDCSGQWRDYDDEGRILVAS 519  
QY 846 TLDPDGTIVNRKVAPEFTDITLDIRVNGISLQGTQFDGSLRLOSQDQ-----G 894  
DB 520 SDPLGQTRRRYDPLG--QILGIELADGSLSYE-YDALGRQTRIDABGSHATLFSKWHG 576  
QY 895 RWNAVYTSAG-----NDQCESTVITPQGFPIHYQYQPE-LDDAVLQVANSNEITQOFSY 946  
DB 577 DILARVSDAGGGLSYLHDBAGRLVALTMENGVQAQFRDILDLRLVBEFTGFDARQRYR 636  
QY 947 NPVTGALLAVNAGGSLPIYVPSGRLKMENIMDKM-----SYMTLARG----- 992  
DB 637 NAADELIARBDDGRETYYAYDRDGLASIRVPATRAPALVRRYKWLMDGRILASAGAD 696  
QY 993 LENGYT-DLTGTIQIKSRDTH--GRVYQIKDS-----SIKTLVYDDLNRHIGSQVTDL-- 1043

Db 697 CEFRATYDEGNL-RLESOVHADQVYVSHSHDALGVRQTSRYGD-----APPVAMLVY 750  
Qy 1044 ATGHM--LTTVEF---DGLNREIGRKLCDSSGHTLDIQOSMLKTOQ-----LAN 1088  
Db 751 GPGHILGALVGAVALAFERDALHREVARDA-----RRDQDALLFTQERQHAFLGRLOR 804  
Qy 1089 RIYKLVAVLORTBOYSYDSNRNL-----NOYKCDGACPTDKYGHST-----V 1131  
Db 805 SRRLAGCPWQRGVYRDGLQGLVGDNDQY-----PSVRYHYDGGGLASRRAGAA 857  
Qy 1132 TQNFYDIYGNITACHTTFADGTEHDATFFKANFTDPCQLTLEVHTHTPMDNIRLKYDK 1191  
Db 858 ASYRYVAAAGN-----RLEGVEFYAR-----EDARQAPAEHELVRSGFSSEVANQ 904  
Qy 1192 AG-----RVINITDNHNTENTFTYDGLRI-----QNGQSVYGYDPLNLYSOKT 1237  
Db 905 AGECPARWAGNRVRIAGN-----RYRFDALGNLVERIGADGERLRLAYDGAQRLVHLTR 959  
Qy 1238 DTLD--CELYRRETMVNEV-----RNG--EMIRLRTGETTIAQGRASKVLLTGDSQ 1287  
Db 960 DYADGTRELEARYRDALSRRIAKVLRDGVQVRFQMDGDRQCAEAPAEHLRTTTHBPG 1019  
Qy 1288 OSVIL-----TSDKONLSQ--EAYSAYGKHKSTANDASITLGYNGRADPVSGVTHLGN 1338  
Db 1020 GFWPLALBQACEBDEPPELQRLQAFABEGPL-----PAQCVPALGE 1062  
Qy 1339 GYRSYDPTLMKFTPTPDISPPGAGINPYSCLDGPIINRSDPSGHLISQWQ---WTGI 1392  
Db 1063 ARIAF-----FHTDH-----LCTPIQLSDERQQLRWQGVDDWRAV 1098

## RESULT 15

US-09-328-352-5486  
Sequence 5486, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 5486  
LENGTH: 3290  
TYPE: PR1  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-5486

Query Match 2.5%; Score 222.5; DB 2; Length 3290;  
Best Local Similarity 18.8%; Pred. No. 5.3e-07;  
Matches 370; Conservative 249; Mismatches 656; Indels 697; Gaps 101;  
Qy 12 ITMSDNNF-----FTQANNFTSAVSG-----GVDPRT 39  
Db 459 LTLSLADNTGVSANDQITQDNKFNKLKGQETGQVTVLSTDESKTMOETTVAKOLAD 518  
Qy 40 GLVNIQITLGHVNGN-----LGPPLPLT--LSYSPINKTIDIGFIGNFGLSYVD 89  
Db 519 GIYQKAVTVDAVAGNISETAIQKVVVDTTAPQAGKLTLSDINDT-----GVSATD 568  
Qy 90 ---RKS-----LISLSTGE-----NYKVIETDKTVKQOKKLDLAREKOLAKENCVRII 136  
Db 569 QITDONSFTLKVQPIVIGEQALLDHYEVSDESKTMOETTVAD---QKDLADGIYQ-- 622  
Qy 137 HKSGLIEVLTGFNNNAFLKVPKLLNPAGHAIIYIDMFEATQPRINRIYDDLGDHDIPL 196  
Db 623 YKA---VITDLAGNISSESAIQKVVVDS-----LVESTTVIVKPIITD----- 663  
Qy 197 LNLVYGLIKITLTLFPQCKEGYRTFLRPLNRQLNSIHNFSLGNNP---LIMSFGYTP 252  
Db 664 ---NTISIVEKQVVISIRLEIANLPTDLS---SLTSVNTTLEAVTYVNFHDE 710

Qy 253 IGRNGLIGOMIT-----SMTAPGGLKETVANSNNNGHHFPOSAN 292  
Db 711 VTO-----EWVTEIRPAFLMSVPEPQTNISIDISLTOAGNTALITHTON---YNVDPTR 762  
Qy 293 LPVLPLYVTLMKQVPGAGQPAIQAEYSYTHNHYGGSSNGIWNKLDNLGLMTBYNGST 352  
Db 763 SPTLDSLT--FNNIDG-----LISGSAKSGKVIIYNKGDWMASTIT----- 804  
Qy 353 ESRRYDKGKH-----DQIVRIERTYNNYHLLTSCKQONGVI--QTTEAYAYAI-- 400  
Db 805 ---NEBKFTLQDLSINTQEVAVANTYNGY-----SSENSIGLVTEVPALISIT 851  
Qy 401 ---IGHNFDSPQPOFOLPKTKETWRSADNSYSEIETTFPDESSE----- 442  
Db 852 RISPEVIGSYATEGSHFIV-----KDQGNILIOEPNSNVFDSGIRFPFVMAUGEY 903  
Qy 443 ---NELTKYIKDKTKQIKIISP-----STHWEY---YPPA-----GEVD 474  
Db 904 RPFILSDQPL-----EEGAQIITSTDKNISGHPQITADYPAVFLBTPQFDSIGETL 958  
Qy 475 NC--PEPPIGYTRFVK---KIOTPPYDSEFKDDEKFIQVR--YSLISQSHVTLKIEBRHY 529  
Db 959 SYHVNENPNSIRAFSGEGNLIATGF-----TDQGFASLQVFPQLKEGFTVQVVDKQ 1013  
Qy 530 SATQL-----NSTLFOYNTDKSELGRLLKQTECTKGE 562  
Db 1014 NTSEBTILEVNFAPVIPHVERITQEGLSGVAENSTIYVADAGNELGKV----- 1063  
Qy 563 NGKTVSVHKEFTYKODDTLOQSHSTTHDNFTIHSQVRSRTG-----R 608  
Db 1064 ---TLGDD---NSWSPFSHPSLVNRPILDEKISVQIINDKGLMSPEQN 1107  
Qy 609 LPSDPTKQIVTQMSYDKGRLLTRTLNSGTPYANTLTVD-----YEANLQDD--- 657  
Db 1108 IIVDLPPAPTEINFEDAGDLV---YGAHPFSKILVKDGGNILKMMWNTDSSGS 1164  
Qy 658 ---NRPPFVITTDVANGN-----OLRN-----EFDG-----AGRHV 685  
Db 1165 FSIHELGFNLNAETVYVATLADVANGNSLAAQIAPNPAFAPYVDSPTSDGIVISQALNNS 1224  
Qy 686 SOCLKSDSGDKFTYTHITQYDEQGRHISTYSYDYLTNQOQDTPDKVHLSMSKSYNNWG 745  
Db 1225 TLVVKDAGD---VVAEIKYGEIDNGWNGSSYFK--LQDLRPLVDGEGFPLSIK--DARG 1276  
Qy 746 QIANTHSYGVSEKLTVDPIITLT--ATKQLQSNNSNNVQTG--KEVYTYPPSQOPIQITLFE 803  
Db 1277 QVS-----ADVTYVADTVAPTPASMLVFSDEGSYTLGVAEIANT-----TIQVFDH 1321  
Qy 804 AGHLQSC--HTLTEDGMDRVRKETDAIGQCTIY-----QYDNYNRVLIQ 844  
Db 1322 NGQLVNIMNNTIIND-----GTFITILYSNNLHGAEATVTVKQAGNVSEALS 1369  
Qy 845 ITLPDGTIVNRKAPFSTDTLIDIRVNGISLQOQ-----TFDGLSRITQODGGRV 896  
Db 1370 INAPLDDI---APNPIKNIILLD--ANGQNFQAQANSAQIEVFDISLG---NOTG--- 1415  
Qy 897 WATYVSAQNDQCPSTVITPDGQFTH--YOQPELDAVLQVANSN-----ITQCFSEYN 947  
Db 1416 WGSTDSAGN-----ASGPFQTYLHGEELTFVVIDRAGNRSIERKONALIDTIAPN 1466  
Qy 948 PV-----TGALLKAVAEQSLTPYYPGSRLLMENINDMKNSYLTMTAGLENGYTDL 1000  
Db 1467 PIENIIFENNGQSTFAQAEAGSSLDVLDQGT-----NKIGF-----GYDS 1507  
Qy 1001 TGTIQKISRDT--HG-----RVTQIYDSSIKITLTANTDLNLRHSGQVTDLAT 1045  
Db 1508 SGNVSGYFOQYVTLHGEELTFVVIDRAGNRSLEVAQSA---LNDVAPNPIENIILLD--N 1562  
Qy 1046 GHMLT-----TVEPGLNRE--IGRKLCDSSGHTLD--IQOSWLTQOOLANRIYVLANGVL 1097  
Db 1563 GQNFQAQANSAQIEIKONNGDVVGVSADBSAGNVSGIYVYVHLHGEELTIVVDRAQ-- 1620

```

QY 1098 ORTEOYSYDSRNLNOYK-----CDGACPTDKXYSIVTQNF-----1136
Db 1621 -----NSTEVKQNALIDIDLPNPIENIVLDINGONFTAOAEANTQIEVKNAV 1669
QY 1137 -----YDIYGNITA-CHTTFADGTEDHATFKFAN-----PTDPCQTEVHTHPMPD 1183
Db 1670 EIVSGYVDAGANVSGLYQVYLHGBE--LTFVVVDAGNRSTEVKQNALIDIDLPNPIE 1727
QY 1184 NLRKTKAGRVNITNHTNGTENTFT-----YDTIGRLONGGSYGYDPLNRLVSQKTD 1239
Db 1728 NIL------DANGONFTAOAEANTQIEVKNAAGEVIG-----SGSTDS 1765
QY 1240 L-DCELYRETMLVNEVRNGSEMIRL--RGETTIAQORASKVLGTGDSQGSVILTSD 1295
Db 1766 MGNVSGIFYQVYL-----HGBELTFVVVDAGNRSTEVKQNALIDIDLPNPIENI 1820
QY 1296 KQNLISQEA-YSAYGKHKSTANDASILQYNGERADPVSG-----VTHLGN 1338
Db 1821 GQNFPTAOAEANSKVEKNAAGEVVGSGY-VDSAGNVSGYLNOVYLKGBELTFVVIDQAGN 1879
QY 1339 GYRSYDPTLRFRHTPDSLSPGAGI-----NPYSYCLGDP--INRSDPSGHL-----1384
Db 1880 --RSIEVKQOTAF--LDNTAPENATNLVPSSEDSYLSGMAEPNATIQIFDYGQQLNOMN 1935
QY 1385 --SMQAWTIGWG--TAGLLLTATGMAIAAAGIAAIASTTALAFALSVTSDI 1439
Db 1936 NVNWDGTFNITLNSNTYHGEVFKV---VVDDQGNLSGEVT-----1973
QY 1440 TSIVSGALBDASPKVSSILGVMGMAAGIABSAIKGTKLATHLGAFAEDGENALLKS 1499
Db 1974 ---VKAPLDDIAPVPAASDLVFNEDGSSLSGVAEP-----NTFIQIFDONGQO--MNT 2020
QY 1500 TSESSRIKMGVTSRLDEIVRNEGOVYKDHRSRGYTDNFMKGEOALVHGDKDGFLYHT 1559
Db 2021 WSQSVNADGTFITFFGTYNLHGEFTVI-----VKDLAGNVSEAVSV-----2062
QY 1560 EGNKHNKGFPYTRHTPEQLVDYLDKNNIIVDLTQGD--KPVHLSCYKSSG 1609
Db 2063 -----KAPLDDIAPNPIKNIIVFDANGSFTAOAEANSQIEIFDSFGSQIG 2107

```

Search completed: January 30, 2006, 09:50:44  
 Job time : 46.5112 secs

**THIS PAGE LEFT BLANK**



```

QY 936 ASNEITQOFSNPTVAGLILKAVAGQSITPIYVSGRLKJENINDMKMSYLM-----T 989
Db 264 TAPETAFEDVDKTSARLLEATNPQGRITTYDVHNOJTGETMDNL--LGGAMETTRQSS 321
QY 990 LRGJENGYTD-----TGTIOKISNDTHGRVTOIKDSSIKTLLNYVDLNRHISQVTDLA 1044
Db 322 LLGRPIKRTDLKKGGAAGAEIRYVDYDGLRIRFINQSLRFTIIDYDVLGQCKVATBDLQ 381
QY 1045 TGHMLTTVEEDGLNREIGRLCDBSSGHTLDIOGSLKTCQLANRIYKLANVLORTQYS 1104
Db 382 AGTGVIIDMEYDDQOEILRTQTJASNOAALTLTQWAVDGLKTRDLQOQSSPLLHETFS 441
QY 1105 YDSNRRLNQYKCDGACEPTDKXGHSIVYQNFYVDIYGVITACHTTFADGTEDHATPFKAN 1164
Db 442 YDPRKRLTLVNYLSSLPDEIQEMTKQIISFDELDNITLCQIRFTDGTSEBAFAKGS 501
QY 1165 P-----TDPQCLTEVHHTHPMPDNIILKYDKAGRVINITDNHGTENFTYDVLGRL-- 1216
Db 502 PGDDHKRCQCLLSIAYTPPRKTPDPTFSYDANGN--QKDEHGN--SLHYDSQSRLIQV 557
QY 1217 -QNGQS--SVYGYDPLNLVYQKTDYDLCEL--YREIMLVNVRANGMIMLLRTGETII 1271
Db 558 AETGAPISQYRYDDHNOLVATR--DGNSEILRFEGHQLSSTVOEDORTYHLHGEQPL 616
QY 1272 AQQ--RASKVLLTGTDSQGSVILTSDKONLSQEAYSYVGHKHSYANDASILGYNGERAD 1328
Db 617 GQGIYVDDEQTLILLITDANQSVMSFGQGLRKAVYSYVGRHSEALSLAGTRNGEARE 676
QY 1329 PVSQVTHLNGYRSYDPTLMRPHYDPSLSPGAGGINPYSCLDGPIINRSPSGH----- 1383
Db 677 AANGVYLLNGYRAYNPPLMRHFSDFLSPFEGGVNPPYCLGPIILRDPGTGHDSGQ 736
QY 1384 -----LSWQ-----AMGIDMGIA-----GILLTLATNGMIAAAGI----- 1416
Db 737 TGLRRPEBGALPMQGGGDIMGWVGEGIVFTVLVGAALITATILGITPYTGPTVULGI 796
QY 1417 -----AAAIASSTYALAFG-----ALSVSDITSIVSGALEDSAPKASILGWVS----- 1462
Db 797 SMTASAAAANVSIVSGALIVGTALNPASTNTANTVAIVNN-----DQTAGVQWMLGIAAV 851
QY 1463 ---MGMAAGLAESAINGKTYLA-----THGAPAEDEGENALLKTSSESER 1505
Db 852 PVGLVFGAGVAVARAIVAAAAKVAANAGTICVRSVSRIQ--LAAAGARRITSSAASAR 909

RESULT 2
US-10-369-493-11243
/ Sequence 11243, Application US/10369493
/ Publication No. US2003023675A1
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO. 11243
/ LENGTH: 1317
/ TYPE: PRT
/ ORGANISM: Methanosarcina mazei
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)--(1317)
/ OTHER INFORMATION: unsure at all Xaa locations
/ US-10-369-493-11243

```

	Query Match	5.8*	Score	51.7*	DB 4	Length	1317				
	Best Local Similarity	21.2*	Pred.	1.5e-26							
	Matches	334	Conservative	194	Mismatches	505	Indels 541; Gaps 72				
Qy	39	TGLYNIQTTLGHVGNGLGPTLPLTL	---	SYSP	LAKTDIGGIG	---	ENFGLSVYRKNSL 94				
Db	56	TGSFYFGQDLSIPGRG	---	LPLTVSRYSNMDNRSG	LFGSGMTFMYNNKLTVDDNNGN		110				
Qy	95	LSLSTGENYK	---	YLEDTKYKLGQCKLDNRF	EKDLKENCYRIHKS	GDLEVLGTGFNN	151				
Db	111	VTVLGGDHDITTYILNPDG	TSRPLSVDDL	---	IKNSDETYTLTKQOT		157				
Qy	152	AFDLKVPKQLN	---	PAGHAIIYDMNEFATOPRLN	RIYDDLGDHIDPLNLE	YQGLIKTI	208				
Db	158	KYNFSSEGLVNI	VDKNGNQI	---	NFTYTGEOGLTKV	TDASGREL	ILAYDHNG	206			
Qy	209	LTLPEGQKEGRTELRPL	NRLQNLN	SHNSLGNENPLTMS	FGPTPIG	---	KNGLSG	261			
Db	207	---	---	HIISTDPGRVMSYSTDDG	GNLIQCNPLIGKLS		241				
Qy	262	---	WTSMTAPRG	---	LKETVYNSNNQGHFPOSANL	PVLPYVTLMKQVAGAP	311				
Db	242	YTYDENHMTISITDP	RGNHPKNTY	---	DEKGVHISG	---	276				
Qy	312	AIQAEYSYTHSNVYGGSG	NGIMNNKLDNL	YGLMTEYNGSTESRRY	KQKXGHDQIVRLER		371				
Db	277	---	NSLNTATY	---	TFNYS	ENRK	293				
Qy	372	TYNVNHLLTSBCKQOQNG	YIOTTEPAAVYAL	IGHNPDGSGSOPOLP	PKTTEYMR	SADNSYRS	431				
Db	294	---	TTED	---	PGKAKTYIFDEHF	W	313				
Qy	432	EITER	---	TFDESGNPLTKVILKDKTKO	KIISPTHEWYYP	PAGEVDCPEPYG	482				
Db	314	ELNETQNGYTI	SYAVDENGRI	SVTENS	KTTKLA	---	YDANGNIILK	364			
Qy	483	FTRPYKAIQIRPYDSE	FPDPEKPIQ	RVYSILIGSGSHYTLK	IEBHAYATOLNS	---	537				
Db	365	YSK	---	SMYYS	---	KNMLISQTELB	NHKSFEIRDNSNLKSIDALG	406			
Qy	538	---	TLFOYNTDSE	ELRLKQTECTK	---	GENG	---	TSYVNHKE	574		
Db	407	HETVSYD	---	KYQOYVGETD	SNKCATFVSNNNGDQ	TTIDANGKTS	SAFYDYTGAVT	462			
Qy	575	---	YTKODDTLOQSHS	ITTHDNFTHRS	QVRSRYGR	---	FSDDTKOI	618			
Db	463	TKTDAGNR	YFOYDALN	LSITD	---	PMGQTTSTNTML	LEIKVLQMLKVDXOSI	516			
Qy	619	VLOMSYD	GLGRLLT	TTLNSGT	---	PYA	---	NLTJYDVELN	LQDDNRPEFVT	665	
Db	517	LTTLLKIS	LKRMQVEX	LGTMMPLA	ITWFPQ	QOKGHTSIDYDPLNRQ	---	VS	568		
Qy	666	TTDVANG	NLRNE	FPDAGRHVS	---	QCLDSDGDKFY	TIHQOYDE	708			
Db	569	VTNALG	TKTRNKYDA	IGNKISIT	INA	YKSTRYSYN	SLNQLVKTATNAMGK	---	VVRNYDA	625	
Qy	709	QGRHHTSYSD	---	YLTNGROQ	TDPDKYHLSMS	SYDMWQOLANTHMS	GVSEKITT	761			
Db	626	VGNLSIT	DENCHKIN	YGDSILNRQVS	YVDALR	KTRNKYDA	VGNKISIT	INA	YKSTRYS	685	
Qy	762	VDPIT	LTLATKQLQSN	NNVQTSKE	VTWYTPSQOPIQ	ITLFE	AGHLQSCHTLTRDGM	DRV	821		
Db	686	Y	---	NSLN	---	QLVKTADANG	GVVRNYDA	VGNLS	---	745	
Qy	822	RKETDA	IGQCTIYQ	YDNVNRV	IQITL	PDGTVNRK	YAPSTDTL	TDIRVNG	ISLGQTF	881	
Db	716	---	TTDANG	KRTYGVDSLNRQVS	ITNALG	TKTRNKYDA	VGNKISIT	DAMWR	---	LTKYSY	770
Qy	882	DGLSL	TQSDG	SGVMA	VTYSA	SGNDQ	PSVILPPD	GOPIHYQ	---	QPELDAV	932
Db	771	DSLNL	VKATYD	MGVVR	ITTYAVGN	---	LIS	TDAGKHTIDY	YDSDROV	SITDPL	825



Qy	933	LOVANSNETQOFSVAPYVPGALLKAVAEQSLPIIYPSGRLEKMEINIMKKKSYLMTLKG	992
Db	826	GRTARNR-----YDAVGKJISSTDEBDGKTSYGYDVLNRLTKVSPDDQKVSY-----	873
Qy	993	LENGYTDLTGTQIKISRDTHGRVTOIKOSIKITLNYDLDNRHI-----GSQV--TDIA	1044
Db	874	-----NYDAVGNRLTMKDSHGTTAVKYDPLNRLLEVLNPDQKVSATYNNK	918
Qy	1045	TGHNL-----TTTVEPDGLNREIGRKLCDSSGHTLD--IQOSWLTQOOLANRIYKL	1093
Db	919	VGNRVKMTYPDQKTTSSYSYDAVNRLLIG--VIDSDGHITSYSYAKKGNLKTMMTPNGV--	973
Qy	1094	NGVLQRTQYGSYDSNRNLNOYKCCDAECPDTRKYGHSIYQNFYDIYGNITACHTTPADG	1155
Db	974	-----KTB-YSDKANRL-----VELINKTTOVASSKYITLLDAGNRLLKVDQLEAG	1022
Qy	1154	TED-----HATEKFNAPDTPCOQTEVHHHTHPMPDN--IRLKYDKARVINYT-	1199
Db	1021	VEGSDSELKESQLLITTYGY--DKLYRLTKV-----DPSKKTYSYKXDSGNISMTT	1072
Qy	1200	--DNHGNTENFYDPLGRLONGQGSVYGYDPLNLVSOKTDTPLDCLY- YRETMVLNVEYR	1255
Db	1073	NVDIGSTISYKYDAADOLLOSGNISYSYDKKGNIKKRVNSTOFMSYSYDE--ANRLK	1123
Qy	1257	N-----GEMIRLRLT--GETTIAQOASRVLLTGDSQOSVILT	1293
Db	1130	NVSEFVSNTMPKSSYNEFYDDGKRILIKTTNGE---MAOSTKYVLDINSALPOVLTE	1185
Qy	1294	SDKONT-----SQEA-----YSAYGKHKSTANDASIL-----	1320
Db	1186	SDTKNTTCYTYGTDLISMTNSBNAYYYHYDGLGSVRSLSDEKGIITKNTLYLDAFGQYOK	1245
Qy	1321	-----GNGEADPVSQGYTHLQNGCRSRDPLMLRPHTPDSLS--PFGAGINPVSX	1366
Db	1246	EIGTVNDEFRLPTGEOMDEDTGLILYRAR--YDPSVGRITVDIKGRVTTQOSINRYVY	1303
Qy	1370	CLGDPINRSDPSGH 1383	
Db	1304	TTNNPVLVLDLTGY 1317	
RESULT 3			
US-10-156-761-7834			
Sequence 7834, Application US/10156761			
Publication No. US20030119018A1			
GENERAL INFORMATION:			
APPLICANT: OMURA, SATOSHI			
APPLICANT: IKEDA, HARUO			
APPLICANT: ISHIKAWA, JUN			
APPLICANT: HORIKAWA, HIROSHI			
APPLICANT: SHIBA, TADAYOSHI			
APPLICANT: SAKAKI, YOSHIYUKI			
APPLICANT: HATTORI, MASAHIRA			
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES			
FILE REFERENCE: 249-262			
CURRENT FILING DATE: 2002-05-29			
PRIOR APPLICATION NUMBER: US/10/156,761			
PRIOR FILING DATE: 2001-05-30			
PRIOR APPLICATION NUMBER: JP 2001-272697			
PRIOR FILING DATE: 2001-08-02			
NUMBER OF SEQ ID NOS: 15109			
SEQ ID NO 7834			
LENGTH: 2364			
TYPE: PRT			
ORGANISM: Streptomyces avermilticilis			
US-10-156-761-7834			

Query Match	4.7%	Score 420.5;	DB 4;	Length 2364;
Best Local Similarity	21.6%;	Pred. No. 2.4e-19;		
Matches 320;	Conservative 190;	Mismatches 550;	Indels 423;	Gaps 73

[illegible]

```

Db 1608 T-----TTT-----LNKVSERQKDT-----TYTDLAKGLTELEQASGQ 1642
Qy 1097 LQREQSYSTRNL-NQYK-----CDABCEPTDKGHSIYQNTYDIYGNITACHT 1148
Db 1643 TAOSQCELYDQALNTAAYHTTTTGICADKTKTASDKGTAPYQATAYTDLGLOSITN 1702
Qy 1149 TPAAGT-----EDHATFKFANPTDPCOLTEV-HHTHPDMPNIRLKYDKARVI 1196
Db 1703 TDSAGTATLADYLYPGVDDTGTTANANOPHGRIKIDHKTGSTTSSDQLYTDDCTMK 1762
Qy 1197 NITDNHNTENFYDTLGRLO-----NGQGSV---YGYDP---LNLVLSQKT---DTL 1240
Db 1763 QRYV-PGITTIDYWTPOGLEAVKTKSSGSELTRYAVADAGNILVTTPOETVASIDGM 1821
Qy 1241 DCELYEFTMLVNEV-RNGEMIRLLRGTETIIAQARASKYLITGTDQOSVILTSDKQN- 1298
Db 1822 ELRTNGTATVATRVASGATVAMRTTEGTTAN--GKVTYLMADYQASTQALVADASTG 1879
Qy 1299 -LSQEAYSAYGKHKS-TANDASILGYNGERADPVSGVTHLNGYRSDPTLMRFTDPL 1356
Db 1880 ASTRRRTTPRDEBSGLPTCTDNGFLCKTEDTSTGLSL--GARAYDPNLRGLSPDPL 1937
Qy 1357 -SPEGAGINPYSCLDGPINRSDPSGHLISQAWTGIQMGIAQ 1398
Db 1938 ATPAPQNLNAYSATNNPISYDPSG-LCRDIDCGGYVAG 1979

RESULT 4
US-10-156-761-7990
; Sequence 7990, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7990
; LENGTH: 820
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-7990

Query Match 4.4%; Score 395; DB 4; Length 820;
Best Local Similarity 23.8%; Pred. No. 2,7e-18;
Matches 242; Conservative 120; Mismatches 375; Indels 280; Gaps 50;

```

```

Qy 736 SMKSVDNNGOIANTHMSYGVSEKITYVDPITLTA---TKOLQSNNNVOTGKEVITYTPS 792
Db 212 SVAGTGTGGNNVYTT-----NPSAPFCGGFAGRCETKIKMTAKVTT----- 256
Qy 793 QQPIQITLFDKACHLOSCH-----TLTRDGMVRKETAIDAGCTIYOYNNY--RV 842
Db 257 -----TFHDSAGNLDTVPAPLAKITYYDALGRKTKYTVDARGVTVVTTNNRDRRI 311
Qy 843 IQITLP-----DGTIVNRKAPFSDTLTDIRVNOISLGGQTFDGLSRLT--QSQ 891
Db 312 VDTTNRAVEYWDGNGNLQR-----TDSGTETIYD-----FDPLORETTITLQ 356
Qy 892 DGR-VWATYTSAGNDQCESTVITPDGQFIHYOYQBELDAVLQVANSNETQOFSYVPVT 950
Db 357 DSGQTLAY-----TPSGVDVYQ-----DAG-----TVDYTNNEV- 388
Qy 951 GALLKAVAEQSLTPYPSGRLK--MENINDMKRNSYLTWTLNGLNGYTDLTGTIOKIS 1008
Db 389 -----NKLAEKDPAGRVTTYKYNNDVRYTT--TYPG-----GTQOKVD 426
Qy 1009 RDTHGRVTOIKDSIKTT-----LNYDDLNRHIGSOV---TDLATGMLTTTVEFDGL 1058
Db 427 PDNSSRPTKITKTSFGTPEVDLAYSYGYGTSGTBESKIRSSIDNVTG--LKITTYIDGA 484
Qy 485 GRFSYAE--BKGTITLN--SSWLYCYDIAGNL-----TSQGTAGGCGATTYTVND 531
Db 1119 ACEPTDKYGHISIVQNTFYDIGNITACHTPADGTEHDATFKFANPTDPCOLTEVHNTH 1178
Qy 532 AQOLTAQNGS---TTWMSYDKIGNETAGAST--PEGRTAEKW-----TDSQSLTSL- 578
Db 1179 PDMPDNIRLKYDKARINITDNHNTENFTYDTIG--RLONGQGSVGYDPLNRLVSQK 1236
Qy 579 -----TWGK--TYTGOYSTDQSERIRLGGTYHANGPGLAG----- 614
Db 1237 TDTLCELYRETMVNEVANGEMIRLLRGTETIIAQARASKYLITGTDQOSVILTSDK 1296
Qy 615 TSTAGVDYTGFR-----BPGGLTNSMTGGKAYCYVTLDAAGSVIALTDE 658
Db 1297 --QNLSEAYSAYGKHKSANDASILGYN--GERADPVSGVTHLNGYRSDPTLMRFT 1352
Qy 659 TGAQVNTYVAYSPRGVTAJSTSEKRPQYRPAAGYOD--VTGLHYHGARY--YDPNIGFNS 715
Db 1353 PDLSPPEGAGINPYSCLDGPINRSDPSGHLISQAWTGIQMGIAGLLTATGMAIYA 1412
Qy 716 PD--PESQER--NEYLYABGDPVNRIDPGLFESDALDTGSDIFGVVTCLAGVTAAB 771
Db 1413 AGGIAAIASTSTTALAFCALSVTSDITSIVSGALIEDASPRASSILGWSMGMAAG 1469
Qy 772 TGSIAVAAA-----VGVVAGVGSVAGVGLAVGSCALG--GAAG 810

RESULT 5
US-10-156-761-7572
; Sequence 7572, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697

```

PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 7572  
 LENGTH: 1250  
 TYPE: PRT  
 ORGANISM: Streptomyces avermitilis  
 US-10-156-761-7572

Query Match 4.4%; Score 392; DB 4; Length 1250;

Best Local Similarity 21.2%; Pred. No. 8.7e-16; Matches 271; Conservative 187; Mismatches 452; Indels 368; Gaps 66;

254 GKNGILGOWITSMAPGKLTETVYNNQGHHPQ-----SANTPLVPLVTLMKQVPG 307  
 Db GMEVAAQGLRK-TANAABDAEDASRQAKICEKDPVDVATGRVLPQI-----DVSL 113  
 308 AGCPALAEYSYSHNAYVGGSGNGIWNKLDNLY-----GLMTEYNYGSTES 354  
 Db PGQLPLVVKQGFESYRLGWFPGPTWSTLDQRLVDVAVLVGEGDLVLAIPH-----PA 170  
 355 RRYKDEKHQOIVRIETNNYHLLTSECKQNGYIOTTEATAYATIGANFDSQSPQOL 414  
 Db PGVPTLPSHGRWMLDRDGGYTL-----TDQKSGHIR----- 203  
 415 PKTKIETWRSADNSYRSEITETFEBSGNPLTKVIXKQKQKIIISPTHWYYPAGEVD 474  
 Db -----HVVNDSDTLAVLEQLDRNGNWT-----FEY-----DAD 233  
 475 NCP---PEPYGTFTRFKV-----KIOTPYDSEFKDPERK-PIQYRSLIGSOSHTLKEE 526  
 Db GAPPVYLTHSGGYRLRISTEAGRVTAHLAASAVDGGQDLIRYV-----TDGHL----- 283  
 527 RHYSATQLANST--LFOYNTDKSELRLKQTE-----CTKGENGKTSYV 569  
 Db -----TEVNASGRLQYVCD--ELGRITSWTDNDHSFYAYDDEDRCTH--OSGAAGHL 335  
 570 VHKPTVYKOD-----DTLOQSHITTHDNFTIHSQVSRARYGRLESPDPTDIYOTM 622  
 Db RSTPFAVAVDPGTAHTTATVTDISGQTH--YLNR-----RCQYIAELDLAGAVTRY 386  
 336 RSTPFAVAVDPGTAHTTATVTDISGQTH--YLNR-----RCQYIAELDLAGAVTRY 386  
 623 SYDGLHLLTRTLNLSGTPYANTLYDY--ELNMLODDNRPPEVITTTDVNGQLNEPDGA 681  
 Db ORDRYNLLSQT--DPLGHTTSFRYDDAGNLVAARP-----DGBAABAEYNAL 433  
 387 ORDRYNLLSQT--DPLGHTTSFRYDDAGNLVAARP-----DGBAABAEYNAL 433  
 682 GRHVSQCLKSDGCKYTHITQOYDECGRHTSYDYDLNCRQOQDPDKHLSMSKSY 741  
 Db GLPKK--LVNPDG-----TTRQTFDBRG-----NLTSVDPF--GQTFRGY 472  
 434 GLPKK--LVNPDG-----TTRQTFDBRG-----NLTSVDPF--GQTFRGY 472  
 742 DNGQIANTHMSYGVSEKITVD-----PITLT--ATKQLOSNNNVQTEGVTTYPSPQ 793  
 Db DEGGRLTSMTDPLGHTTGIVCDRAGLPVTVDPDLAGAVTRYERDAFGRTA--ITDPTGAT 530  
 473 DEGGRLTSMTDPLGHTTGIVCDRAGLPVTVDPDLAGAVTRYERDAFGRTA--ITDPTGAT 530  
 794 QPITITLFDAGHLQ-----SCHTLTRDGMHVRKETDALIGQCTIYQDYNRVIT 846  
 Db TRLEMTY---EGHLSRRTPARPGTESMTYDEGNCSTHTDPVGVSLFEYTHPLPRTART 587  
 531 TRLEMTY---EGHLSRRTPARPGTESMTYDEGNCSTHTDPVGVSLFEYTHPLPRTART 587  
 847 LPPGTVINRKXAPSTLTITDIRVNGISLQOQTPDGLSRLTOSQDGRWATYSAGND 906  
 Db GPDGV---RY-BFEND--TELR-----LSQVTPH--GLTMVYAYDAAGR 624  
 588 GPDGV---RY-BFEND--TELR-----LSQVTPH--GLTMVYAYDAAGR 624  
 907 QCESTVITPQGFHYOPELDAVLOVAS--NEITQPSY-NPVTGALKKVAEAGSL 963  
 Db LVAETDP--DWRITLYEY-----DPRAGLSRNLALQGMIAFENELGQIVRKDAQOAT 677  
 625 LVAETDP--DWRITLYEY-----DPRAGLSRNLALQGMIAFENELGQIVRKDAQOAT 677  
 964 TPVYPSGRLKMNINDKMSYLMTLRGLNGYTLDTITQIKISRLTHGKVTQIKOSI 1023  
 Db TVANVDFDQLAQA-----TGPDGTALTLIRBHGMSSEAVNGR 716  
 678 TVANVDFDQLAQA-----TGPDGTALTLIRBHGMSSEAVNGR 716  
 1024 KTLIANYDNLNHT-----GGOVT---DLA-----TGHMLTTTVEVDGLANREIGRLCD 1068  
 Db ALTYDYDELGRTRRTTPSSATITMSYDVAGRRTSMTSGSISIDFVUDEAGRELGRRP-- 774  
 717 ALTYDYDELGRTRRTTPSSATITMSYDVAGRRTSMTSGSISIDFVUDEAGRELGRRP-- 774  
 1069 SSGHTLIDQGSWLTQQLANRIVLNGVLTQRTBOYSVDSNRRLNQYCGDABECTDKYGH 1128

Db --GH-TLHSPDALGRITSGSV-LGPAGRRTOHRY-----TYRAD-----GH 814  
 775 --GH-TLHSPDALGRITSGSV-LGPAGRRTOHRY-----TYRAD-----GH 814  
 1129 SIYTON-----FTYDIYGNITACTH-----TF--ADGTEHATFKFANP----- 1165  
 Db LIHIEDQLSGRRFFDLADAGRVTAHAVNWTETAYADAVNGQSASAPAGPQGEALGNR 874  
 815 LIHIEDQLSGRRFFDLADAGRVTAHAVNWTETAYADAVNGQSASAPAGPQGEALGNR 874  
 1166 -----TDFCQLETHHTHHPM-----PNIRLKYDKAGVINITDNHGTEN 1207  
 Db TYTGSLITRAGVRYEIDGGRILILQKRLSKRPDTRWTWAEDELTSVTPDGTWRMR 934  
 875 TYTGSLITRAGVRYEIDGGRILILQKRLSKRPDTRWTWAEDELTSVTPDGTWRMR 934  
 1208 FTYDTLGR-----LQNGGSYGYDPLNRLVSOKTDLDCELYRETLVNEVR---N 1257  
 Db YTDPLGRKRAKRLAEDGETLV-----BRVFTWDTGVLCEQTTSPLPHQVTLTWD 988  
 935 YTDPLGRKRAKRLAEDGETLV-----BRVFTWDTGVLCEQTTSPLPHQVTLTWD 988  
 1258 GEMIRLLRTEGTTIAQGRASKVL-----LVTGDSQQSIVLTSDKNLSQEAAYSAY 1307  
 Db HOGRLPVTQTERIVADAAPQGEIDSREFALVTDLVGPSE---LDERGSIWTRATLW 1045  
 989 HOGRLPVTQTERIVADAAPQGEIDSREFALVTDLVGPSE---LDERGSIWTRATLW 1045  
 1308 GKXKSTANDAST--LGNGSRADPVSGVTHLGNGYSYDPTLRFHTPDSLSPGAG-GI 1364  
 Db GSTTWAKNSTAYVPLRPFGQYDPEGLHY--NYFRHYDEPTARVYTPD--PLGLGPAP 1100  
 1046 GSTTWAKNSTAYVPLRPFGQYDPEGLHY--NYFRHYDEPTARVYTPD--PLGLGPAP 1100  
 1365 NPYSYCLGDPINRSDPSG 1382  
 Db NPAAY-VPNPHMADPLG 1117

# RESULT 6

US-10-156-761-7751  
 Sequence 7751, Application US/10156761  
 Publication No. US20030119018A1  
 GENERAL INFORMATION:  
 APPLICANT: OMURA, SATOSHI  
 APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIKAWA, JUN  
 APPLICANT: HORIKAWA, HIROSHI  
 APPLICANT: SHIBA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHITUKI  
 APPLICANT: HATORI, MASAHIRA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-262  
 CURRENT APPLICATION NUMBER: US/10/156, 761  
 PRIOR FILING DATE: 2002-05-29  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 7751  
 LENGTH: 2386  
 TYPE: PRT  
 ORGANISM: Streptomyces avermitilis  
 US-10-156-761-7751

Query Match 4.4%; Score 391.5; DB 4; Length 2386;  
 Best Local Similarity 20.1%; Pred. No. 2.6e-17; Matches 379; Conservative 214; Mismatches 688; Indels 601; Gaps 93;

163 NP---AGHAYIDMNFETAPRLNRIYDDGDHPIPLNTEYQGLKITILTLPGQKRGY 219  
 Db NPDGCTIKTMLYIQKTL-----YGDGDDVLPVAINNRKIDL-----DNKVG- 740  
 698 NPDGCTIKTMLYIQKTL-----YGDGDDVLPVAINNRKIDL-----DNKVG- 740  
 220 RTBLRFLNRQNSIHNFSLGNENPLTWSPG-----YTPIGKNG 257  
 Db SAEILNF--PRIQIENG-DLGATTKVSYGAFANACIDHLPQASNTODCYQKMTPEBETE 797  
 741 SAEILNF--PRIQIENG-DLGATTKVSYGAFANACIDHLPQASNTODCYQKMTPEBETE 797  
 258 ILGOWITS-NTAAGGLKETVYNSNNQGHFPPSALNPLVPLPYTLMKQVGG-----AGQ 310  
 Db SKTGWPKKFLVTRKVEDPVP--TTNODG-----APVMTTSYTYEGAGWREFTG 844  
 798 SKTGWPKKFLVTRKVEDPVP--TTNODG-----APVMTTSYTYEGAGWREFTG 844  
 311 PAIOAB-----YSYTSNHYVGGSGNGIWNKLDNLYGL-----MTSEYNGSTESRRYKD 359

Db 845 PLVDEBSMDRGQVOTVTTGADTQKTKMYLGRGLDGRSTKIDTSATKTVTVND 904  
 Qy 360 KEGHDQVRIERTNNY-----HLLTSECQONGYIQTTEFAAYAILGN---PDSQ 408  
 Db 905 GGDG-----NNYADHAMLAGHTLSTLSRDTDI--SHEHTYHTYSHNRQYDGL 951  
 Qy 409 PSCQPLKTKETWRSADNSYRSEITETTPDES-----GNPLTKYIKCKTKOKIISPS 461  
 Db 952 PDAFVRESKSTNTTKISGGRBHVENEYDSEASATTEGLPMR---TDMWGSNVS-- 1006  
 Qy 462 THMEYTPPAGEVDN--CPPEPYGTRFPVKKIQTYPDEHFKDDPEKFIQYKSLIGSOSH 520  
 Db 1007 -----DNRC-----TYGGRAYNTD-----NYDSTGAQRWT 1031  
 Qy 521 TLKIEBHYSATQJLNSTLFOYNTD--KSELGRLLKQECTGKNGKTSYVNHFTYTKQ 578  
 Db 1032 VVOQVQVHYVG---CSSIADSNODGYTSTLYDNATSIDANKPVDGNATEY--RITTK- 1084  
 Qy 579 DDTLQOSHSTIYTHDNFTIHSQVRSRYTGLFSDTDTEDIVTQMSYDKLGRLLTRTLNSG 638  
 Db 1085 -----AGRYRSSWTG-----YDAGRIVMS--EDG 1107  
 Qy 639 TPYANTLTVD---YELNNLODDN-----RPPFYI-----TTTDVNGQ 673  
 Db 1108 KHNRSITTYSPANTWPMNGITSTSPDPGTATARGPLTFETWTSRPMATPTTSKDXANGNI 1167  
 Qy 674 LRNFPGAGRHVSQCLKDSGDG---KP-YTHHTQO-----YDEQGRHHTSTYS 718  
 Db 1168 TKVTLDAAGREVEYWKPTETGSSPSMKFSYTIPTSTMSAGVPSDADGYPHVAHSHLQSGS 1227  
 Qy 719 DYLTN-----GR-----OQTPDKVHLMSKS-----YDNMGQIA----- 748  
 Db 1228 TYLVSHAVTDGLGARARETQPLPSVDVPAKQIVPFRQVAVTRDSAGQVAGASVFRNQ 1287  
 Qy 749 NTHMSYGV-----EKITVDPITLTKQLOSNNSNVQYCKEYTY-----TP 791  
 Db 1288 GTASGGSPSSPOPBDLPSYSDLVDMAGRIVSSQIQVNGTPQKAGRVDTSLGDTYSVTP 1347  
 Qy 792 SQQPIQITLFPDAG-----HLOSCHTLTRDGW---DRVKEKEDPAIQCTIYQYDYNR 841  
 Db 1348 VDSRTTDTTYTVYQVSKVVEHTASAVTTAYGTAKELAQITDPKGNNTLYYDMAAQ 1407  
 Qy 842 VIOITLPPDGTIVNRY---APFSTDTLITDIRVNGISLGOQTFGLSLRLOSQDGC---R 895  
 Db 1408 RKTDDPAGAGSSSEYENNGQVSOQTATNTDVQVLTG---YDNLSTATSVRSGADELA 1464  
 Qy 896 VWAY-----TYSAGNDQCPRSVITPDDQFIHYQPELDDAVLQVANSNETQQ 943  
 Db 1465 AMVWDDPAATGKQGITSAVSRDASGNITTKTKP-----DERGRPLMTTVTLR 1514  
 Qy 944 PSYNPVTGALLKAVA--EGOSLTPYYP--SGRLKXENI---NPMKKMSYMTLRLGLENG 996  
 Db 1515 TYVNGLAGDYTTSVYYDAADHITSYSPAGKLAEXKTYTYDDYGOPTRLTSSIG--- 1570  
 Qy 997 YTDLTGTIOKISRDTGRVTOIKDSSIKTYLN-----YDILNRHISGO-VTDLATG 1046  
 Db 1571 ---GTAVIDNTTYAVGRALVE--RDVGAFFGGNGIOAQROGYDDBSN---GTRMLRSAT- 1622  
 Qy 1047 HMLTTFVFDGLNREIGRKLCDSSCHTLDIQOSMLKTOQLANRYKNGVLQOREQIYSYD 1106  
 Db 1623 ---TTINDLVSEAKOTYLLVDNTGKLELRB-----QASGQTAQSGCLREYD 1666  
 Qy 1107 SRNRLN-QYK-----CDGABCPPTDKYGHISIVTQNFYDIYGNITAGHTTFADDT----- 1154  
 Db 1667 DOSRLTLAYHTTGAACADTTKTTSDFKGTSPYQGYTYDRLGNQSVTIDNSAGAATTR 1726  
 Qy 1155 -----EDHATFKFANPTDPCQLTEVHHTHPMDNIRLKXDK-----AGRVINIT 1199  
 Db 1727 DVLVGYDODAGTWTAN-----ADQHGVR-KINKVSAGTTAAAGTYTYA 1771  
 Qy 1200 D-----NHGNTFETFDYTLGRL-----QNGQGSVYGYDPLANRVSQKITDLDLDEL 1244  
 Db 1772 DGAMKQRVESGTTTDTYTWSRLGRLATVTKTSGSGLTRRYDDASGNLLVETTP----- 1825

Qy 1245 YRETMV---NEVRNGEMIRLRT-----GETTIAQORA-----SKVLLTGTDSQOS 1289  
 Db 1826 --GETVASIGTLELRTDGSATATRYSGATTVAMRFTDGNNTVNGKITLMLGDTQAS 1883  
 Qy 1290 VILTSQKQ--NLQOAVSAYGKHS--TANDASILTNGEADPVSGYTHLGNRYSTDP 1346  
 Db 1884 TOIADVAAAGTAAFRRRRTTPGDERSGSLPTGYNHGFGKTEDNTGSL--GARADPS 1941  
 Qy 1347 LMFHTPDSLS--PGAAGINPYSCGDPIINRSDPSGHLMSOA-----WTGIGMGA 1397  
 Db 1942 LGRFLSPDLSTPYDPONLSAISYSGNDPIINYSDPSGLILKNSDGTQCSGWMQKCGEYV 2001  
 Qy 1398 G-----LL-----LTATGMAIAAG-GIAA 1418  
 Db 2002 GGGGDTPTPAFVQDDLVLDLPREBNGWADRLAQVWHYQGTTOGGGYWDAPVGDGDR 2061  
 Qy 1419 AIASTSTAL--APGALSVTSDITSIV-----SGALEDASPAASSILG 1459  
 Db 2062 GMACFGRTACSEAFVWKETHDPAKARVATPCVENPKRCGADNGAYDSMKESAEPVI 2121  
 Qy 1460 WBSMGAGLAASAIKGTKLATHLGAFADGENALLK--STSESRIKG-----V 1510  
 Db 2122 LLAGEMGA-----FSKVLKARCCSFRTTRVLMDGKTKPLKIKRPGDLVEADP 2172  
 Qy 1511 TRSLDREI-----VRNEGQVTKDS---RGYTDNFMKGEOALIV 1548  
 Db 2173 TSGHREVRVAVTAVLHNDODLVLSIRGLDGRILQTLHTTARHRIMDTAQVMEQARLI 2232  
 Qy 1549 HGDKDGFYHTEGNKH-----NGKPYTHHTPEQLVDYLDKNNIVLDTQGGDPR-- 1557  
 Db 2233 TGKH-----VNTSGOHATITSVLAORGAADMDLYVGLHTY-----YVLAGEPVL 2280  
 Qy 1598 VHLSCYKGS-----GAADKMAKYINRPVIAVSNKPTISQGLARIERKDFLKSTYHS 1651  
 Db 2281 VHNGSCWSTNRKTSVKNAGHMKH-----KSEFPNLMNAKEYIABGTDPLRST--- 2330  
 Qy 1652 YDPKIIIGRTKTEKVPKPTRP 1673  
 Db 2331 -DPS-VLTRTRANGDVIRFNP 2349

RESULT 7  
 US-10-282-122A-46565  
 : Sequence 46565, Application US/10282122A  
 : Publication No. US20040029129A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Wang, Liangsu  
 : APPLICANT: Zamudio, Carlos  
 : APPLICANT: Malone, Cheryl  
 : APPLICANT: Haebelbeck, Robert  
 : APPLICANT: Ohlsen, Karl  
 : APPLICANT: Zyekind, Judith  
 : APPLICANT: Wall, Daniel  
 : APPLICANT: Trawick, John  
 : APPLICANT: Carr, Grant  
 : APPLICANT: Yamamoto, Robert  
 : APPLICANT: Forsyth, R.  
 : APPLICANT: Xu, H.  
 : TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 : FILE REFERENCE: ELITRA.034A  
 : CURRENT APPLICATION NUMBER: US/10/282.122A  
 : CURRENT FILING DATE: 2003-02-20  
 : PRIOR APPLICATION NUMBER: 60/191,078  
 : PRIOR FILING DATE: 2000-03-21  
 : PRIOR APPLICATION NUMBER: 60/206,848  
 : PRIOR FILING DATE: 2000-05-23  
 : PRIOR APPLICATION NUMBER: 60/207,727  
 : PRIOR FILING DATE: 2000-05-26  
 : PRIOR APPLICATION NUMBER: 60/230,335  
 : PRIOR FILING DATE: 2000-09-06  
 : PRIOR APPLICATION NUMBER: 60/230,347  
 : PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 46565  
LENGTH: 2234  
TYPE: PR1  
ORGANISM: Bacillus anthracis  
US-10-282-122A-46565

Query Match 4.3%; Score 386; DB 4; Length 2234;  
Best Local Similarity 18.7%; Pred. No. 5.7e-17;  
Matches 333; Conservative 236; Mismatches 633; Indels 608; Gaps 76;

QY 21 FTGANNFTSAVSGGVDR-----TGLYNIQITLGHIVG-----NGNL-----57  
DB 748 YSKVGHASEISHFVRFSGDSGFIMPFYMASIPVLNGKVNATNGNFI MSEKDITLSG 807  
QY 58 -GPTLPITLSYSPANKTDIGFGIGFNGLSV---YDRKNSLISL-----TGEN-102  
DB 808 RGDVSVERTYNSQSKVGLFGTGWSSGLEERVWADNGNLLITDGANITFRTGDNK 867  
QY 103 -----YKVIETDKTYKLGQKKLDNRFKDKLKNCTYRIHSGDEVL 145  
DB 868 YQAPFTGYLEIKOVSGSEYKIKDQYV-----TFYSGDQGR 905  
QY 146 TGRNNAFDLKP-----KLLNPAGHAIYIDNFEATOPRLNRIYDDLGHDIPL 196  
DB 906 IETKQKYGTTTYYEYDGAIRLSKVNASKELVL---QYDGNKKARV-----GRPKT 959  
QY 197 LMEYQGLITITLPPGQ---KSGYRTFLRLNQLNSINPGLNENPLTMSFGY---250  
DB 960 ITENYGDLLVSSITTEGKYKGYD-----NGVLTSIYDPOHTDAKPKTSYAYENDR 1013  
QY 251 -----PIGKNGIL-----GQWITSMTAPGLKELTVNNSNNQGHHPQSANLVLRYVTL 301  
DB 1014 LVKVTPELGAATTLANTYTSKEVTLINPKGRKTVYTTND-----AGNPP-----1057  
QY 302 MKQVPAGQPAIOAEYSYTSNHYVG-----GNGIWMNKLDMLYGLMTEYN- 348  
DB 1058 -KTVEDVGRNLTTSTVEYNNANLVKTTTPKNGQETELATYDNGNVTSTVDEMGTBEKPYNK 1116  
QY 349 -YSTSRRKRYKE-----GHDQVIERITNNTHLLTSECKQONGYIQTETAYYAI 400  
DB 1117 DNDIIRKATDNEDRKTTVAAYVANTVEVSQTDGAN-----TSSVYIHDQYGNPIETSKLS 1171  
QY 401 IGHNPDSQSPOLPKTKETETWRSADNSVASEITETTFDSSGNPLTKVIXDKTKQKILSP 460  
DB 1172 AGGNLIONBS-FEM--NGTEKWKVVDNNSGSIS---KDATPAPGGLGSSSLKITT 1223  
QY 461 STH--WEYYPAGEVNDNCPREPYGTFRFVKII-----QTPYSEFPDPP 503  
DB 1224 ATNNDGXYIAAIEVLEPWTYTLTSGMYKTDLVNGAAPPVOSLWNGAGIDGCHMDR 1283  
QY 504 EKRIQYRISLIGSQSHVTLKIBERHYSATQLNLSLTFQYNTDKSELGRL-LKQTECTKGE 562  
DB 1284 HNKVQSTSDVMNQ--VTFKTEQ---TRKVIYIQLVENGSGATSGSAMFDFIOLEKGE 1337  
QY 563 NGKTYVVMHFTYTKQ--DDTLQO-SHSITTH-----DNFTIRSOVRSYTGRLTS 611  
DB 1338 VSSSRPVLANSFEBNMPDGFVPQWARSQSCQCBRDVSDSTGHSIYMER-----S 1391  
QY 612 DTDTKDIIVTQMSYDKLGRLLTTRLNNGTPYANTLTYDYELANLQDDNRPFPVITTDVNG 671

DB 1392 EYGPNDI-----GYRRVILNQGKATVTLTAMSKSENVND-----1428  
QY 672 NQLRNEFDAGRHVSQCLKDSGDGKFTYTHIQDYDQGHHTSTSYDYLTNGROQDP- 730  
DB 1429 -----APDKLSKDYAVLAETYYODG-----TVNNYYSFSPSTNDW 1464  
QY 731 -----DKVHLSMSKSYDMWQGLANTHWSGVSEKITYDPDITLPAKQLOSNS 777  
DB 1465 NRSNAYIPAKPKIQKIEIFLFRKNNG-----KWFDDIRLBNALIKNE 1511  
QY 778 NNVTGKEVTTYTPSOQPIQITTFDEAGHLQSCHTLTRDGMDBRKETDAIGQCTIYQYD 837  
DB 1512 YD-NDGNVAVTY-----DEEGCK--NTFTYDASGNKKSSTDEKNTKLYDYN 1555  
QY 838 NNRKVIQITLPDGTYNRKAPSTDTLTDINVNGISLQOQTFDGLSLRLOSODGGRW 897  
DB 1556 KDMILTRKVTLNKSTSVNRYD-----1576  
QY 898 AYTYSAGNDQCPSTVLTIPDQGFVHYQYQPELDVAVLQVANSNETTQGFSPNPTGALLKAV 957  
DB 1577 -----HNGNTTKSVMFQGKTQTHKREYDVNKNVTYIDALNRRIEN-TYDENANKIKTKM 1631  
QY 958 ABQSILTPYTPSGRLMENINDMKKWSYLMTLRGLENGYTDLTGTLQKISRDPTHGRVQ 1017  
DB 1632 PNGSILLESYVDTADRVGK-----RNGKDSFT-----FERDQNGVTK 1670  
QY 1018 IKQ--SISIKTLTYDLNRIHISQVTDLATGHLTTTVEFDGLNREIGRKLCDSSGHTLD 1075  
DB 1671 VKDLVNGVERTKYTDVADR-----VTS-AT-----DSRGKID 1702  
QY 1076 IQQSWL-----KTQGLANRYKLVNGVQLRTQYQYDSNRRLNO--YKCDGACPTDKY 1126  
DB 1703 -----MAYHDKANSKTEKLEQYVTOGQY--TNKVSYD-YNTLDQNI RVTDSQ-----1748  
QY 1127 GHSIVTQNFYTDYGNITACHTTFADGTEDHATFEK--ANP-----TDPCLTEVYHT 1177  
DB 1749 -----TYRFDYDQGNV-----RTYTAGNSGSGTFNPDQANKIDLVGTSINSLSRYE 1799  
QY 1178 HPMPDNIRLKVDKAG-----RVINITDNHGTENFTYDTLGR-----LQ 1217  
DB 1800 YDQSGNRKTKHKGAGAGKVTETNFVYDPINQILNEVLPNQTTKSYTDGFGNRTSVKIE 1859  
QY 1218 NGQGS---YGVDPRLRLVSQKTDLDCEI-----1244  
DB 1860 NGEKTSIATTEBEGQLVYFGNESLTYDVNGNRSTDGKKYTWNEDDQIVAITKQGENN 1919  
QY 1245 -----YREBTLVNEVRNGEMIRLRTGETI-----IAOQ 1274  
DB 1920 AFATYKYDDBDNRIEKNVAGQVTRYFYDDDSINPLVETDNGTVLQYVVSADGRLAMK 1979  
QY 1275 RASKVILLTGTDSQOQVY--LTSDKONLSQEAAYSAYK-----HKSTANDASILGNGERA 1327  
DB 1980 AQCQTLTYHYNPRGVDVAMTNOCKEVAVATYEDAMGNVLTSDTKGIAAD--NPGYAGYMY 2038  
QY 1328 DPVSGYTHLGNNGRSVDPFLMRHTPD-----SLSPFAG--GINPYSYCLGSPINR 1377  
DB 2039 DKEIGMYTL-----IARYNEBHGVLFSVDDPDEDEDDPVTMNGTYVADNNPVM 2088  
QY 1378 SDPSGHLISQWAMTGMIGMAGLILLTATGGMALIAAGSIAAIASTSTYTLAAGALSPTS 1437  
DB 2089 TDPDGMAM-----LVPVVAGAMVLAARFQAKYAL-----RYGAKTGKKAIVKS 2131  
QY 1438 DITSIVSGALDEASPASSILGWSNMGAGLAESAIKGYTLATHLGAFADGENALL 1497  
DB 2132 -----GWDY-----GKVVAKSGMKKGSIAOKIPRHHVGR---I 2163  
QY 1498 KSTSESSRLKMGVTRSLDBREIVANNEGOYIKDHSRYTNNPFGKQDALIVHDDKQGFV 1557  
DB 2164 KQDNDGKGYGVYISYTK-----KTKGRTYS-----SFEFHTPHNGKGY 2203  
QY 1558 HTEGNKNG--KGPYTR 1572  
DB 2204 HLQKNKYSKYQGMNR 2219



```

APPLICANT: Yamamoto, Robert
APPLICANT: Foreych, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITPA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27,931
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 76214
LENGTH: 843
TYPE: PRF
ORGANISM: Salmonella typhi
US-10-282-122A-76214

Query Match 4.0%; Score 353.5; DB 4; Length 843;
Best Local Similarity 22.8%; Pred. No. 2.3e-15;
Matches 223; Conservative 121; Mismatches 308; Indels 327; Gaps 48;

QY 542 YNTDKSELG---LTKQTECKENGKTVSVHAFYTKODDTLQSHSITTHDNFTIH 597
DB 32 HHTDSGQYRLDMVLAERSLCVTDSMGRT-----RCHWMDAGGLVAY 74
QY 598 RSQVRSRYTGR-----LPSDPTTKOIVTOMSDKXGLLRTLTNGSTPANTLTYYE 650
DB 75 RDEAGQMTTFWSDERLLGMTDAQGGKMYVVDRLGHL-----TETHD-P 120
QY 651 LNNLODNNRPPF---VITTTDVNGNQLRNEFD-----GAGRHSOCL 689
DB 121 LGRVEQOMHFWHQPTEDVDAAGVAMRYEYDERGNLQANSDDLQRTVYGYDRH-GQVV 179
QY 690 K--DSDDGKYYTIHQYDEQG--RH-----HTSYSDYLTNGHQDTDPDKVHLSMS 738
DB 180 RITDARGGDKL-----QWNEGDQLMHTDSCSGQTAMFDERRLRLRYDABE--NSTR 232
QY 739 KSYVNMGOIANTHMSYVSEKITVDPTLTATKOLQSNNSNVQCKEYTTTPEQPIQI 798
DB 233 YSYVGNHLETFWFADEGRTERYQPD-----AAGRIVKYSFAG--QI 272
QY 799 TLPEAGHLGCHTLTRDGMVRKETAIGCTIYOYDNNRYIOTLPGSTIVNRKYA 856
DB 273 TRWQ-----RDGGRVRRQTDATRRRTAYEDAYGRLLTILNGBESYRFRY- 319
QY 859 PFSFTDLITDIRVNGISLGQOTPDGLSRLTQSD--GGRVMAVYTSAGNDQCP----- 909
DB 320 -----DVLRVKTEQTGPGGSRAYGYVGNALNAVAVAYGGERG 356
QY 910 -----STVITPDGQFIHQYQPELDDAVLYQ-----ASREITQOF 944
DB 357 GEIRHGLERDAAGRLTAKITPE---TRREYVYDADARLLEIRRRHDAEGGEDEVI-RF 412
QY 945 SYNFVTGALLKAVAG-----QSLTPIYPSGRLLKMNINDMKMSYLTMLRGLN 995

```

```

DB 413 SYNSAGNLSEFAAGVYLQHRVDVQGNRTETQMPDGR-----TIRLXY 456
QY 996 GYTDLTGTCTICKISRDTHGRVTOIKDSSIKTTLAYDNLNHSQVDTLATGMLTTTVER 1055
DB 457 G-----SGHLQOI-----NIGRVDISEFTF----- 476
QY 1056 DGLNREIGRLCDSGGHTLDIQSWLKTQOLANRIV--KXNGYLQRP---EAYSYSRNR 1110
DB 477 DHLREYOR---SQGN-LMRRMYDPTGRLTRLCTCKGRGVVPEFTIREYAYSQDE 531
QY 1111 LNOYKCDGACBPDYKXSHSVTONFTYDIYGNITAC-HTFPAGTEBDHATFKFANPTDPC 1169
DB 532 LTK-----KRHSRQGVTDYF-YDTTGRITACGNEAYLDSWQDAA---ANLIDR 577
QY 1170 Q-----LTEVHHHPDMPDNIRLTKDKAGRVINITDNGNTENFTYDTL 1213
DB 578 QGETAQAAGSVVPEFNRITSYRGLH-----YRYEYGRVVEKRGKRG--TQHYRWDAA 628
QY 1214 GRL-----ONGGGSVYG--YDPLNRLYSOKTDTLQCELYRRETMVNRNGEMTILR 1265
DB 629 HRLTEVAVTGRTVRRGYVDAPGRREVERHELDABEKFPNRTFTLWDGRLAQECLGR 688
QY 1266 TGETIQAQR-----ASKVLITGTD-SQOSVILTSQKMLSQEA--YSAY 1307
DB 689 SSSIXIYSDRSGHEPLARVDRAAPGEADRYLHYHTDVNGAPREMTDGGNIWEAGYQW 748
QY 1308 GK--HKSTANDASILGNGERADPVSGVTHLANGYRSYDPTLMRFTPTDLSLSPFG-AGG 1363
DB 749 GNLTHEKETRPVQONLRFQCGYLDREFGHY--NLVYFYPDIGKFTISGD---PIGAGG 803
QY 1364 INPYSYGLDPIRNSPDSG 1382
DB 804 INLYQYA-PNPLSTYDPLG 821

RESULT 10
US-10-282-122A-47600
Sequence 47600, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Maloney, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreych, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITPA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22

```







```

Db 218 RTQTFHREAAEFGSGITGVLT-DGAGRHFLVLVLTQORAEARQAISGTEBSAPDPT 276
Qy 462 -----THMEYYPAGEVNCPEP---YGF-----RFVK 488
Db 277 LPTGYEGRNGIRLSAVMLTHDPEYD---ENLPAAPLVRYGWPBRELAVYDQSNST 331
Qy 489 KIITQPYDSEFK-----DDEKFIQYRYSLSIGSHTVTKIERHYSATQLNLTSL 539
Db 332 QVRSFTYDDKXRGMAVHRTGRPE--ICRYD---SDGRVTBQLNAGLSYT----- 379
Qy 540 FQYVTDSELGRLLKQTEC--TKENGKTVSVHKKFTYTKODDTLQOSSHITTHDNFTIH 597
Db 380 YQEKDRIITLDSLNREBVLTHQEGG-LKKVVK-----BHADSQV 421
Qy 598 RSQVRSYTGRLPSDPTKQIVTQMSYDKGRLLTRTLNLSGTPYANTLTYVLELNNQDD 657
Db 422 QSQPDA--VGRKQOTDAGRTTYSPPVYGLTIRLT--TPGGRASAFY----- 468
Qy 658 NRPEVITTTDVNGNQLRNEFDGAGRHVSQCLKSD-----GDGKFTYIHTQYDEQGRH 712
Db 469 NHHSQLSATSAPDGLAIRREYDEMGRLIOETAPDGLITRYKYNPHBDLPATBDATGSR 528
Qy 713 HTSTSYDLTNGRQOTDQKXVHLSMSKSYDNMGQIANTHMSYVSEKITVDPI-TLPAVK 771
Db 529 KTWMSRY---GQLLSFTDCSGVTRYDHRFGQVTAHREBGLSQRAYDSRQGLIAYK 585
Qy 772 QLOSNANVQNGKEVTTTPSQPIQITLTPDEAGHLOSCHTLRDMGDRKEDATGQ- 830
Db 586 DTQCHETRYE-----YMAADLTT--VIADPG--SRNGTQYDANGKA 623
Qy 831 -CTI-----YQYDYNRVLIQTLPDGTVINRKAPFSTDTLITDRVANGISLGQOTF- 881
Db 624 ICTTGGGLTSMEDVAGRVIRLISENGS-----HTFRVDVDRLLQETGFPDRTQRYH 678
Qy 882 -DGLSRLTQSGODGRVNAVYTSAGNDQCESTVITPDQFIHYQYQPELDAVLQVANSNEI 940
Db 679 HDLTKGRLRSEDEGLV-----THMYD-BADRTHRTVANGET 714
Qy 941 TQGSVNPVGCAL--LKAVALGOSLTPY-YPB-GRLKMEI-----NDM--KMSY 986
Db 715 AERQOYB-RGWLTDHISHISGRVTVHYGDSGRILASELTVNHPQTEMLMOSHTRH 773
Qy 987 LMTLRLNGEYVDTLGTIKISRDTHGRVTVQIKOSIKTTLVNDLNRHIGSOVTDLATG 1046
Db 774 AYNAQGLAN--RCITPDSLPAVEMLTYG-----SGMLSCKMG 808
Qy 1047 HMLTTFYEF--DGLNREIGRKLCDSSGHTLDIQSMWLTQOLANRIVKLVGLORTEQYS 1104
Db 809 D--TPLVEYTRDRLHRETLR---SFGRYELTAYTPAQLOSG--HLSNLS-DRDYT 858
Qy 1105 YDSNRNLNOKYCDACEPTDKYGSIVTQNFYIYGNITACHTTPAD-----GTEDHA 1158
Db 859 WNDNGELIR-----ISSPRQ-----TRSYSSTRLRGTGVHTAANLDIRIPYTDPA 906
Qy 1159 TPKFANPTDPCQLTEVHHTHPD-----MPDN-----IRLKYDKAGRVNITD----- 1200
Db 907 GNRLPDP-----ELHPDLSLWMPDRNIRAKDAHYLYRYDRHGLTEKIDLLEBEGV 956
Qy 1201 ---NHGNTENFYDTLGRLONGOSVYG-----YDPLNRLVSKOTDLDCEL----- 1244
Db 957 IRTDDEKTHRYHVSQRVLVYTRQYAPLVBESRYLYDPLGRVAVARWRERENDLGM 1016
Qy 1245 ---Y--RETMVNE-----VANGEMIRLR---TGETTIAQORA--- 1276
Db 1017 SLSRKPQVTVYMGDGLTITQNDRTRIQTIYQGSFTPLIRVETAGELAKTORSLAD 1076
Qy 1277 ---SKVLLGTDSQOSVILTSDKOMISOA----- 1303
Db 1077 TLOQSGEDSGSVFPVLVQMLDRLESEILA---DRVSESRWMLASCGITVAVQMSQM 1133
Qy 1304 ---YSAVGRKXSTANDAS-----ILQYNG 1324
Db 1134 DPVYTPARKIHLVYCHDRGLPLALISTEGTTAVYAEVDWGNLNLNENPHQOLILRLPG 1193

```

```

Qy 1325 ERADPVSGVTHLNGVRSYDPTLMRFTPLPSLSPFG-AGGINPYXCLADPINRSDPSG 1382
Db 1194 QQIDBSGLVY--NRHRYDPLQGRYITTD---PIGLKGMNFIQYPL-NPISNIDPLG 1246

RESULT 12
US-10-282-43060
: Sequence 43060, Application US/10282122A
: Publication No. US20040029129A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl
: APPLICANT: Zykkind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.
: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: ELITRA.034A
: CURRENT APPLICATION NUMBER: US/10/282,122A
: CURRENT FILING DATE: 2003-02-20
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/230,335
: PRIOR FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: 60/230,347
: PRIOR FILING DATE: 2000-09-09
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/267,636
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 78614
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 43060
: LENGTH: 1411
: TYPE: PRT
: ORGANISM: Escherichia coli
: US-10-282-122A-43060

Query Match 3.8%; Score 340.5; DB 4; Length 1411;
Best Local Similarity 20.4%; Pred. No. 4.3e-14;
Matches 287; Conservative 172; Mismatches 459; Indels 490; Gaps 70;

Qy 418 KTEFW-RSADNSRSRSETETTPDESGNPLFKV-----KDKTKOKIS-----PS----- 461
Db 218 RTQTFHREAAEFGSGITGVLT-DGAGRHFLVLVLTQORAEARQAISGTEBSAPDPT 276
Qy 462 -----THMEYYPAGEVNCPEP---YGF-----RFVK 488
Db 277 LPTGYEGRNGIRLSAVMLTHDPEYD---ENLPAAPLVRYGWPBRELAVYDQSNST 331
Qy 489 KIITQPYDSEFK-----DDEKFIQYRYSLSIGSHTVTKIERHYSATQLNLTSL 539
Db 332 QVRSFTYDDKXRGMAVHRTGRPE--ICRYD---SDGRVTBQLNAGLSYT----- 379
Qy 540 FQYVTDSELGRLLKQTEC--TKENGKTVSVHKKFTYTKODDTLQOSSHITTHDNFTIH 597

```

Db 380 YQEKDRTITDSDREVLHTQCBAG-LKRVVK-----EHADSVT 421  
Qy 598 RSQVRSKRTGRLFSDDTKDLYVQMSYDKLGRLLTRTLNSTGYPAANTLYDELANNO 657  
Db 422 QSQPDA--VGRRAQOTAAAGRTTESPDVVYGLTRIT--TPGASAFYNNHNO-- 473  
Qy 658 NRPFVITTTDVNCSQJNEFPDAGRHVSQCLKSD-----GDGKFTIHTQYDEGRH 712  
Db 474 ----LTSATPDDLELRREYDELGRLIQETAPDGDITRYVNDPNSLPATDAGSR 528  
Qy 713 HTSTSYDYLINGROQTPDKVHLMSKSYDNWGOJANTHMSYVSEKIVDPI-TLPAVK 771  
Db 529 KTMWMSRY---GQLSFTDCSGVYTRVDHDFGQMTAVHREBGSQYRAYDSRQGLIAVK 585  
Qy 772 QLOGNSNNVQTKGKVTYTPSQRPQITLFEBAHQ-----SCHTLTDGMDRVRKE 824  
Db 586 DTQGHETRYE-----VNIAGDLTAIVAPDGSRNQTOYDAMGKAVRT 626  
Qy 825 TDAIGQCTIYQDYNVRYIOTLPDGTIVNRKVAPESTDTLITDIRVNGISLGQOTF--D 882  
Db 627 TQG-GLTRSMETDAAGVIRLISENGS---HTFRYVLDRLIQTGFPDGRTORVHND 680  
Qy 883 GLSRLOSQDGRWAWYYSAGNDQCPSTVITPDQPIHYOPELDAVLQVANSBITQ 942  
Db 681 LTGLLISEDEGLV-----THMYD-EADRLTHRYVKGETAE 716  
Qy 943 QPSTNPVTGAL--LKAVALGOSLPIIY--PSGRL-KMENINDMKMSYLM----- 988  
Db 717 RMQYDE-RGWLTDISHISEGRVA-VHYRYDEKSLTGEROTVHNPQTEALLMOHETRA 774  
Qy 989 -TLGLENGYTDLTGTLKISRDPHGRVTOIKDSIKTLLAVYDLNRIIGSOVDTLATGH 1047  
Db 775 YNAQGLAN--RCIPDSIPAVEWLYTG-----SGYLAGMKLGD----- 809  
Qy 1048 MLTTTVEF--DGLNRLEIGRKLDSGHTLDIQOSMLKTQOLANRIVKLVGLQRTQESY 1105  
Db 810 --TLPEVETRLRHRLETLR-----SFGRYELTAVTPAGQLOSO--HNLNLSL-DRDYTW 859  
Qy 1106 DSRRLNIOYKCDGACFPDCKYGHSLVTONFTYDIYGNITACHTTPADGETEHAFFKPNP 1165  
Db 860 NDNGELLR-----ISSPRO-----TRSYSTGTGRLTGVHTAA-----NLDIRIPYA 902  
Qy 1166 TDPC--OLTEVHNTPD-----MPDN-----IRLKYDKAGRYVNTD-----NH 1202  
Db 903 TDPKGNRPD--PELHPDSTLSMMPDNRIARDAHYLYRDRGRLEKTDLIPBGYIRTD 961  
Qy 1203 GNTENFTYDTLGRLONGO-----GSVYGYDPLNRLVSOQTDLLDCEU----- 1244  
Db 962 ERTIRYHNDQHRLLVHTRTYEERPLVBSRYLDPGRVAKRVRMRERDLGMMSLSRK 1021  
Qy 1245 ---Y---RETMVNE-----VANGEMIRLR---TGETTIAQQA----- 1276  
Db 1022 POUVWYMGDGRLLTIIQNDRSRIQTIYQPSFTPLIRVETATGELAKTORSLADALQOS 1081  
Qy 1277 -----SKVLLTGTDSSQSVTLTSDKNLSQEA----- 1303  
Db 1082 GGBDGSVVPFVPLQMLRLSEILA--DRVSESRMTLASCGLTVEQMKQMDPYT 1138  
Qy 1304 -----YSAYGKHKSSTAND-----ASILGYNGERADP 1329  
Db 1139 PARKIHLVCHORGLPLALLISTEGATWCAEYDEMGNLNENPHQLOQLRLPQOQYDE 1198  
Qy 1330 VSGVTHLNGKRSYDPTLMRFTPLSLPFG-AGSINYSYCLGPINRSPSGHLSMOA 1388  
Db 1199 ESGLYV--NRHRYDPLQGRYITQD---PIGLKGMNLYGYOL-NPISIDIPILGSMWED 1252  
Qy 1389 WTGIGMGIALGLLTATGGMALAAAGGIAAALASTTALAAGALSVSDITSISYSGLE 1448  
Db 1253 -----AKSG--ACTNGHCGTISA-----MIPDPKFDSDSTAY 1283  
Qy 1449 DASPRASSILGVNWSMGMAAGLAEBSAIGKGTGLATHLGAFADGNNALLKSTSESSRIKW 1508

Db 1284 DALINKINS-----QSIDCKEPA----- 1301  
Qy 1509 GYVRSUDREIVRNEBGQVTKDHSRGYTDNFMKGEO-----ALLVHGKDGFLYHTEG 1561  
Db 1302 -----GLICDNGSRIFSTLAPNNGERKGSYFPNPPCNGTEKVSAYHTHG 1346  
Qy 1562 NKHNGK--GPYTRHTPEOLVDYLDKNNI 1587  
Db 1347 ADSHGEYWDIEIFSGKDEKIYK-SKDNMI 1373  
  
RESULT 13  
US-09-912-020-340  
Sequence 340, Application US/09912020  
Patent No. US20020045592A1  
GENERAL INFORMATION:  
APPLICANT: Zykkind, Judith  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Trawick, John  
APPLICANT: Forsyth, R. Allyn  
APPLICANT: Froelich, Jamie M.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
FILE REFERENCE: ELITRA.001DV1  
CURRENT APPLICATION NUMBER: US/09/912.020  
CURRENT FILING DATE: 2001-07-23  
PRIOR APPLICATION NUMBER: 09/492,709  
PRIOR FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: 60/117,405  
PRIOR FILING DATE: 1999-01-27  
NUMBER OF SEQ ID NOS: 485  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 340  
LENGTH: 1426  
TYPE: PRN  
ORGANISM: B. Coli  
US-09-912-020-340  
  
Query Match 3.8%; Score 339; DB 3; Length 1426;  
Best Local Similarity 20.2%; Pred. No. 5.5e-14;  
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;  
  
Qy 572 KFTYTKDDTLQOSHSTTHDNFTIHSQVRSKRTGRLFSDDTKDLYVQMSYDKLGRLL 631  
Db 317 RYTYTEAGELL---AVYDRSNTQVRAFTYDAQHPGMVAVHRYVAGREMRRYDDTGRV 372  
Qy 632 TRTLNSGTPYANTLTLYEYELNLODDNRPPFVITTTD----- 668  
Db 373 EQLPAG-----LSYTY---LYEQDR---ITVDSLNRREVLHTEBGAGLKRVAKE 418  
Qy 669 -VNGNOLRNEPDGAGRHVSQCLKSDSDGKFTYHTHQYD-----BGRHHTSYSDYL 721  
Db 419 LADSVRSRGVDAAGRLTAQ--TDAAGRTEYGLNVVSGDITDITPDGRRTKFTYND-- 474  
Qy 722 TNGEQCT--DPRVHLMSKSYDNWGOJANTHMSYVSEKIVDPIITLTKLOQNSN 778  
Db 475 --GNQLTAVVSPD--GLSREYDEPERLV-----SE 502  
Qy 779 NVQSGKEVT--TYTPSQRPQITLFEBAHQSGCHTLTRDGMVRYKETAIGQCTIYQD 837  
Db 503 TNSGSEVRYRYDDAHSELPAITTTDANG--STRQMTWSRYGQLLATDSCGYOTRIEYD 559  
Qy 838 NVNNEVIOITLPDGTIVNRKVAPESTDTLITDIRVNGISLGQOTPDGLSRLOSQDQ-GRV 896  
Db 560 RFGM-----TAVRE-----EGSL-YRAYDNGRLTSVDAQGRE 595  
Qy 897 WATYSGANDQCPSTVITPDQPIHYOQ----- 925  
Db 596 TRYRYNNAAGDL--TAVITPDGNGRSETQYDAMGKAVSTTQGLTRSMEXDAAGVYSLTNE 653

```

Qy 926 -----BELDAVLQVANSNEITQOFSYNPVTGALLKAVAGOSLTPYPSGRKME 976
Db 654 NGSHSVSYDALRLVQGGFDRGTQRYHD-LTGKLTQSEDEGLVLTWYDESDRITHR 712
Qy 977 NINDMKKSYLWTLRGLNGEYDTLGTIOKISPDTHGRVTOIKDS-----IKTTLANDL 1032
Db 713 TVNGEPARQOY-----DGHGMLTDISHLSGHRVAHVYGDdk 751
Qy 1033 NRHIG--SOYTDLATGML----- 1049
Db 752 GRLTGCQYVENPEPTGLWQHETKAVYNOGLANRVTPDLPPEVWLYTSGYLAKMKL 811
Qy 1050 --TTTVEF--DGLNREIGRKLCDSSGH--TLDIOQSWLTKOQLANR----- 1089
Db 812 GCTPLVEYTRDLRHLRETVRSGMAGNAVELSTTPACQLOSLNSLYVDRDYGMS 871
Qy 1090 ----YVLNGVLQRTQOYSYDSRNRLNQYKCDGAC-----PTDKYHSHI----- 1130
Db 872 DNGDLVHISGPRQ-TREYVSATGRLESVRLAPDLDIRIPYATDPAGNRLPDELPDHS 930
Qy 1131 -----VTQN-----FTYDIYGNITACHTTPADG--TEDATFPKANPTBCQLTEV 1174
Db 931 TLTVPMNRKLAEDAHVYRDEYGRLEKTDRLPAGVIRTDDEBT----- 975
Qy 1175 HHTHPDMDNRLKYDKAGRVINITD-NHGN--TENFTYDTLGR----- 1215
Db 976 HHNH-----YDSQHLVFYTRIQHGEPLVESKYLDPDLGRRAKAVMRERDLTG 1025
Qy 1216 ----LQNGQSVYGYDPLNRLVSOQTDLTDLCELYR----- 1247
Db 1026 WMSLSRKEVTVYGMDC-DRLLTVQTDTRIQTVYEGSGFTPLIRVETENGEREKQGRS 1084
Qy 1248 --ETMLVNRVANG-----EMIRLL-RTGETTIA-----OQRAKY 1279
Db 1085 LAETLQOEGSENGHGVVFPALVELLDRLEELRADVSSBRAMLAQCGLYVEQLARQV 1144
Qy 1280 LLAGTDSQOS-----VILTSDKONLSQEA-VSAYGKHKSTANDASILGYN----- 1323
Db 1145 EPEETTPARKAHLHCHDRGLPALLISDGTAMSALEDEMGONLEENPHHV-YQPYRL 1202
Qy 1324 -GERADPVSQVTHLNGVRSYDPTLMRFHPDLSPPG-AGGINPVSYCLADPINSRPS 1381
Db 1203 PGQOHDEBSGLYU--NHRRYDPLQGRYITQD--FMGLKGWMLYQYPL-NPLQOQDPM 1256
Qy 1382 GHLSQWMTGIGAGIAGLLTATGMAIAAGIAAIAASTSTTALAFGLASTSDITS 1441
Db 1257 GLL--QTMWDARSG-----ACTGV--CGVLSRIIGSPKFDSTADALD----- 1296
Qy 1442 IVSGALDASPKASSILGWYSMGWGAAGLASAIKGGTKLATLHGAFADEGENALLKSTS 1501
Db 1297 ----ALKEQNRS-----LQNDMEYSIGVCKDTNG-----KYFASKAETDNLK-K 1336
Qy 1502 ESSRIKMGVTRSLDRLEIVRNEBGOVIXDHSRGYTDNFMGKGBQAILVHGDK--DGLLYHT 1559
Db 1337 ESYPLKCKCPTGTDRVAAYHTHG--ADSHGDYVDEFFSSSKVLVSKONNLEAFYLAT 1393
Qy 1560 EGNKH--NGKPY 1570
Db 1394 PDGRFELANNKGEY 1407

```

## RESULT 14

```

US-10-282-122A-42617
Sequence 42617, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeelbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel

```

```

APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 42617
LENGTH: 1426
TYPE: PR
ORGANISM: Escherichia coli
US-10-282-122A-42617
Query Match 3.8%; Score 339; DB 4; Length 1426;
Best Local Similarity 20.2%; Pred. No. 5.5e-14;
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;
Qy 572 KFTYKQDDTLQOSSHITTHDNFTIHSQVRSYRGTLPEDTDTKOIVTQMSYDKGRLL 631
Db 317 RYTYEAGELD-----AVYDSRNTQVRAFTYDAQHPGMAHRYAGRPENRYYRDTGRVV 372
Qy 632 TRTLNAGTPYANTLTVDYELANLQDNRPEFVITTD----- 668
Db 373 EQLNPAAG-----LSYRY-----LYEQDR-----ITVTDLSLRREVLTBEGAGLKRVYKKE 418
Qy 669 -VNGNLRNEPFGAGRHVSQCLKDSQDGGKFTYTHQOYD-----EQGHHTSTSYDYL 721
Db 419 IAGGSYTRSGYDAAGRLTQ--TDAAGRRTYGLANVSGDITDITPPDGETRYVND-- 474
Qy 722 TNGRQOT--DPPKHLSMSKSYDNMGQJANTHMSYGVSEKTIYVDPITLPAKQLQNSNN 778
Db 475 --GNQLTAVVSPD--GLSRRREYDEPERLV-----SE 502
Qy 779 NVQTKREV-TYTPSOQPIQITLFDAGHLQSCHTLRLRDMVRKETAIGQCTIYOYD 837
Db 503 TSSGFTVRVRYDDASELPATTTDAG--STRQMTWSRYGGLAFPTDQSGVQTRYEYD 559
Qy 838 NNRKRVQITLPDQTYNRKAPSTDTLTLDKRVNGISLQOQTFDGLSRLOSQDQ-GRV 896
Db 560 RFGOM-----TAVHRE-----EGISL-YRYEDNRGRILTSYKDAQRE 595
Qy 897 WATYSAGNDQCSYVTPDQCFIHYOY----- 925
Db 596 TRYENAAADL-TAVITPDGARSFTQYDAMKCAVSTTQGLTRSMETDAAGVISTLNE 653
Qy 926 -----BELDAVLQVANSNEITQOFSYNPVTGALLKAVAGOSLTPYPSGRKME 976
Db 654 NGSHSVSYDALRLVQGGFDRGTQRYHD-LTGKLTQSEDEGLVLTWYDESDRITHR 712

```

Qy 977 NINDMKKMSYLMTLRLGLENGYTDLTGTLQKISRDPHGRVYQIKDSS-----IKTLANTDDL 1032  
Db 713 TVNGEPABQWQY-----DGHGMLTIDISHLSGHRVAHYGYDDK 751  
Qy 1033 NRHIG--SOVYDLATGHTL----- 1049  
Db 752 GRLTGBEQYTBENPRTGELLMOHETKHAVNEQGLANRVTPDLPPEVWLTYSGYLAQMKL 811  
Qy 1050 --TTTVEF--DGLNREIGRLCDSSGH--TLIDIOQSWLKTQOLANR----- 1089  
Db 812 GGTPLVEYTRDLRLRETVRSFGWAGSMAAYELSTYTPAQLOQSHLNSLVYDRDYGWS 871  
Qy 1090 ----IVKLVGLQRTBOYSYDSRNRNLNQYKCDGAE-----PTDKKGHSI----- 1130  
Db 872 DNGDLVRISGPRQ--TREYGSATGRLESVTRLAPDLIRIYATDPAGNRLPDELIAPDS 930  
Qy 1131 ----VTQN----FTYDIYGNITACHTTPADG--TEDHATFKFANPTDPCQLTEV 1174  
Db 931 TLTVWPNRILAEADAHVYRHDYGRLTTEKTRIPAGVIRTDERT----- 975  
Qy 1175 HHTHPMDNIRLKYDKAGRVINITD--NHGN--TENFTYDTLGR----- 1215  
Db 976 HHYH-----YDSQHRLVFYTRIQHGEPVLESRYLXPLGRMAKRVWRERBDLTG 1025  
Qy 1216 ----LQNGQGSVGYDPLNRLVSOQKTDLTDCELYR----- 1247  
Db 1026 WMSLSRKEPVYTWGMDG--DRLTIVQDTTRIQTYEBSFTPLIRVETENGEREKAQRSS 1084  
Qy 1248 --ETMLVNEVNG-----EMIRL--RTGETTIA-----QORASKV 1279  
Db 1085 LAETLQOEGSENGHGVFPALVRLDLREIRADRVSSERAMLAQCGLTVEQLAROV 1144  
Qy 1280 LLTGTDSQGS-----VILTSQKUNISOA--YSAVKGKHSSTANDASILGN--- 1323  
Db 1145 EPEYTPAKKALYHCDHRGLPLALISEBDNTAMSEYDENQUNEEPHV--YQPYRL 1202  
Qy 1324 --GERADPVGTVTHGNGYRSYDPTLMRPHTPDLSLSPFG--ACGIPYSYCLGDPINRSDPS 1381  
Db 1203 PGQGHDESSGLY--NHRKYIDPLQGRYITOD--PMLKKGWMLYQVPL--NPLQQLDPM 1256  
Qy 1382 GHTLSQWANTGIGWIGAGLLTITATGMAIAAGGIAAIASTSTTALAFGALSVTSDITS 1441  
Db 1257 GLT--QTWDDARSG-----ACTGCV-----CGVLSRIIGPSKFDSTADALD----- 1296  
Qy 1442 IVSALBEPASPRASSIILGVSMGMAAGLAESAIGKTKALTHLGAPEDEENMLIKSTS 1501  
Db 1297 ----ALKETQNRS-----LCNDMEYSIGVCKDTNG-----KYFASKAETDNLK--K 1336  
Qy 1502 ESSRIKMGVTRSLDEIVRNEEGVYIKDHSRGYTDNFMKGEOAILVHGDK--DQFLYHT 1559  
Db 1337 ESYPLKRCPTGTDRVAAYHTHG--ADSHGDYDEFTSSSDKNLVRSKNNLEAFYIAT 1393  
Qy 1560 EGNKHA--NGKGPY 1570  
Db 1394 PDGRPEALNNKGEY 1407

RESULT 15  
US-10-771-241-340  
; Sequence 340, Application US/10771241  
; Publication No. US20040241715A1  
; GENERAL INFORMATION:  
; APPLICANT: zykend, Judith  
; APPLICANT: Foreyth, R. Allyn  
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
; FILE REFERENCE: ELITRA.001C1  
; CURRENT FILING DATE: 2004-02-03  
; CURRENT APPLICATION NUMBER: US/10/771,241  
; PRIOR APPLICATION NUMBER: 09/492,709  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/117,405

; PRIOR FILING DATE: 1999-01-27  
; NUMBER OF SEQ. ID NOS: 485  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ. ID NO 340  
; LENGTH: 1426  
; TYPE: PRT  
; ORGANISM: E. Coli  
US-10-771-241-340

Query Match 3.84; Score 339; DB 5; Length 1426;  
Best Local Similarity 20.24; Pred. No. 5.5e-14;  
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;

Qy 572 KFTYTKODDTLQOSSHITTDNFTIHSQVRSRYTGILFSDTDTKDIVTQMSYDKLGRLL 631  
Db 317 RYTYTTEGELL--AVYDSNSTQVAFATYDADHPGMAHRAVGRBEMRYRDDTRRV 372  
Qy 632 TRTLNPGTAYANTLYDYELNLTQDDNRPPEVITTD----- 668  
Db 373 EQLNPAQ--LSYRY--LYEQDR--ITVTSILNRREVLHTEGAGLKRIVVKE 418  
Qy 669 --VANGNOLRNEFDGAGRVISQCLKSDGDKFYIIHQVD-----EGGRHHTSYSDYL 721  
Db 419 LAQGSVTRSGYDAAGRLTAQ--TDAAGRTEYGLNVVSGDITDTPDGRBETKYND-- 474  
Qy 722 TNGRQOT--DPDKVHLMSKSYDNWGQIANHWSYGVSEKITVDPIITLATKQLQJNSN 778  
Db 475 --GNQLTAVVSPD--GLSRSREYDEPERLV-----SE 502  
Qy 779 NVQGEKVT--TYTPSQOPIQITLFEAGHLQSCHTLTRDGMDRVREKETAIGQCTIYQD 837  
Db 503 TSSGSETVRYRDAHSELBATTTDAG--STROMTWSRYGOLLAFTDCSGYQTRYEYD 559  
Qy 838 NNRKRVQITLPDGIYNRKXAPSTDTLTDIVNGISLQOQFPDGSRLTQSGDS--GRV 896  
Db 560 RFGQM-----TAVRE-----EGIST--YRRYDNRGRLSVDAQRE 595  
Qy 897 WAAVYSAGNDQCESTVTPDGOFLHYQY----- 925  
Db 596 TRYEYNAAGL--TAVITTPGNSSETOYDAMGKAVSTTQGGLTRSMEMYDAAGRVISLTNE 653  
Qy 926 ----BELDAVLQVANSNETQOFSYNPVYTGALLKAVAGOSLTPYPSGRLEME 976  
Db 654 NGSHSVSYDALRVLVQGGFGDGTGRYHYD--LTGKLTQSEDEGLVLWYDESDRITHR 712  
Qy 977 NINDMKKMSYLMTLRLGLENGYTDLTGTLQKISRDPHGRVYQIKDSS-----IKTLANTDDL 1032  
Db 713 TVNGEPABQWQY-----DGHGMLTIDISHLSGHRVAHYGYDDK 751  
Qy 1033 NRHIG--SOVYDLATGHTL----- 1049  
Db 752 GRLTGBEQYTBENPRTGELLMOHETKHAVNEQGLANRVTPDLPPEVWLTYSGYLAQMKL 811  
Qy 1050 --TTTVEF--DGLNREIGRLCDSSGH--TLIDIOQSWLKTQOLANR----- 1089  
Db 812 GGTPLVEYTRDLRLRETVRSFGWAGSMAAYELSTYTPAQLOQSHLNSLVYDRDYGWS 871  
Qy 1090 ----IVKLVGLQRTBOYSYDSRNRNLNQYKCDGAE-----PTDKKGHSI----- 1130  
Db 872 DNGDLVRISGPRQ--TREYGSATGRLESVTRLAPDLIRIYATDPAGNRLPDELIAPDS 930  
Qy 1131 ----VTQN----FTYDIYGNITACHTTPADG--TEDHATFKFANPTDPCQLTEV 1174  
Db 931 TLTVWPNRILAEADAHVYRHDYGRLTTEKTRIPAGVIRTDERT----- 975  
Qy 1175 HHTHPMDNIRLKYDKAGRVINITD--NHGN--TENFTYDTLGR----- 1215  
Db 976 HHYH-----YDSQHRLVFYTRIQHGEPVLESRYLXPLGRMAKRVWRERBDLTG 1025  
Qy 1216 ----LQNGQGSVGYDPLNRLVSOQKTDLTDCELYR----- 1247  
Db 1026 WMSLSRKEPVYTWGMDG--DRLTIVQDTTRIQTYEBSFTPLIRVETENGEREKAQRSS 1084



**THIS PAGE LEFT BLANK**

Db 1302 EATLMSPKGMAY-----DKNGLIYFVDG--TMIRKVDNGIISTLL-----GSND 1344

```

Qy 218 GYRTELAFNLRLQNSIHNFSGENPLTWSPGYTPIGKNGILQOMITSMATPAGLKETVN 277
Db 1345 -----LTSARPLTCD---TSMHISOVRLEMPDL-ALNPMDSIY 1380
Qy 278 YSNNQGHHPQSANLEVLPRVYTLMKQVPG---AGQPAIOAESYTSNHYVGGGNGIY 333
Db 1381 VLDNNVLAQITENQVRIAGRPWHQCPGVGEYEVGKHAQVT---TLESAPALAVSYSGV- 1437
Qy 334 NMLKLDNIYGLMTEVNGSTESRRYKDKEGHQIYRIERTYNNHLLHSEKQON----- 387
Db 1438 -----LY---ITETDEKXINRIRQVTTDGEISLVA-----GISECCCKNDANCDC 1480
Qy 388 -----GYIQTTE-----TAYVALIGH-----NF-----D 406
Db 1481 YQSGDGIKAKAKSLAPSSILASPDGTLYIADLGNIRIRAVSKNPKELNSNFYVVASPTD 1540
Qy 407 SQPSQFOLPKTKETETWSADNSY-----RSBITETPEDESGNPLTKYIKD---KTKQK 456
Db 1541 QELYIFPINGHQYVSLVGTGDLNFSYNDNDITAVT---DSNGNTL-RIRDEPNRMPVR 1598
Qy 457 IISPSSTH---WEYYPAGEVUNCPEPYGFTFVKKIIQTPYDSRFPKDDPEKFIQYRSLI 514
Db 1599 VVSPDNQVITMLTIGTNGCLK-----GMTAQGLEVLVFTYH----- 1633
Qy 515 GSQGHVTLKIEBRHYSATQLNSTLFOYNTDKSELGRLKOTECT-----KGNKGKTYIS 568
Db 1634 GNSGLATKSDERQW-----TTFPOVDSR---GRLTNTVFTPTGVNTLHGMMDALIT 1682
Qy 569 VVHKFYTKQDDTLQOQSHSITTHDNF---TIHRSQVRSRYTGLFSODTDKQIVTQMSYDKL 627
Db 1683 VDISSSSREEDVSI---TSNLSISIDSFYTMVQDOLRNSY-----QIGYDGS 1725
Qy 628 GRLLTRTINSTPYANLTL---YDYEINMLQDDNRPPV---ITTDVNGQLRNEPFGAG 682
Db 1726 LRIT-----TASGLDSHYQTEPHVLAGTANPYAKRNMPLPGENGQWL----- 1768
Qy 683 RHVSQCLKSDGDGKFYTIHTQOYDEQGRHHTSTYSYDYLINGROOTPDVKHLSMSKSYD 742
Db 1769 --VEMRPKEQAQK-VVVFGRKLRVNGRNLISVDPRTTTEKIKYDHRKFL-LRLAYD 1824
Qy 743 NMGOIANTHWSYGVSEKIVTDPITLTATKQLOSNNSNVQSEKVTYTPSQPIQITLFD 802
Db 1825 TSGH---PTLML---PSSKLMANVNTYSSGTQIAS---IQRG-----TTSERK---VDYD 1866
Qy 803 EAGHLQSHLTLRQG---WDRVRKEDAI-----GQCTIYODNVRVQITLTP----- 849
Db 1867 GQGRIVS---RVFADGKTWSYTYLESKMWLLHSGQRYIFETDMMDRLSALIMPVSARHTM 1924
Qy 850 -----GTIVRKYAPFSTDTLITDIRVNGISLGOQTFDGLSRLTQSGDGRVWATYISA 903
Db 1925 QTIHISIGYRNINVPESNASIITDYNEGILL-QTAFIAGTSR-----RV----- 1968
Qy 904 GNDQCPSTVITPDGOFIHYQOPELDDAVLOVASNEITQOQSYNRYVTGALKAV---AEG 960
Db 1969 -----LFFKRRQTRLS---ELIYDSTVFSFYDETAG-VIKTVLQSDG 2008
Qy 961 QSLFPIYPSGRLKMKENINDMKMSYUWLTGLJENGUYDLGTGICKISRDHGRYTOXLD 1020
Db 2009 FICTIRYQIQLPL-----IDRQIFRPSSEDMVNAARD-----YIDNSFRYTSQOG 2054
Qy 1021 SSIKTTL-----NYDDLNRHIGS-----QVTLATGMLTTFVEFD---GLNREIGR 1064
Db 2055 VINERPLRIDLQFPDIDSGKVGQFGVYIYDINOIISTAVMTTKGFADHGRKELQY 2114
Qy 1065 KLCDSGGHTLDIQOSWMLTKQOLANRIYKLVGLQRTQOYSDSRNRRLNOYKCDGACSTPD 1124
Db 2115 EIFFELMWTITQYD---NMGRVTKREIKI-GPFAVTTKAY-----EYVDVG-QLOQT- 2162
Qy 1125 KYGHSIVVQNFYVDIYGNITACHTTFADQTEDEHATFKFANPRPQCLTEVHHHTHDMDN 1184
Db 2163 YVLNKKIMRWYNDYDNGNL-----HLNPNNSARLTP----- 2194
Qy 1185 IRLKDYKAGRVINITD-----NHGNTENFTYDTLGRLO-----NGQGSVGYD 1227

```

```

Db 2195 --LRVLDLRITRLBGVOVRLDEDFLRORGTEIFEYSSKGLLTRVYSKSGWTVIYRYD 2252
Qy 1228 PLNRVLSQCT-----DTLDELYY---RETMLVNEVANGE 1259
Db 2253 GLGRRVSSKTSIGLQHLQFPYADLITPTPLTHVYNNHSSSETSLYDLOGLHPLAMEISSGD 2312
Qy 1260 MIRLRTGETIILAQBPASKVLVTGTDQOQSVILITSDKONLSQBPASAYGK--HKSTANDA 1317
Db 2313 -----EFLYASDN-----TGT---PLAVFSSNGMLKQIQTAYAGEIYFDSNIDFQ 2355
Qy 1318 SILGNGERADPYSGVTHLGNRYSDPTLMRHTD-----SLSPFG----- 1360
Db 2356 LVIGFPGGLYDPLTKLIHFGE--RDYDILAGRWTTPDIEIWKRIKQDPAPFNLMPERNNN 2413
Qy 1361 -----AGGINPYCYLGDPIRSD---PSGHL-SWQAWTGIGMGI 1396
Db 2414 PASKHVDKDYITDVNSWLVTFPGLHNLIPGFVPKPFDTLBSYELVKSQOQDDI-PPI 2472
Qy 1397 AGLLTIATGMAIAAAGIAAIAAS-----TSTTALAFLALSVTSD---I 1439
Db 2473 FGVQGVVARGAKAFLSIGKMAEVOVRRRAGAQSWLMWFAVKSILGKVMIAVSQGRVQ 2532
Qy 1440 TSIVSGALEDPASPRASSIIG-----WVSMGMAAGLAESAIIKGTK 1480
Db 2533 TNYVLIANMEDCI-KVAIVLNNAFYLENHFTIEGKQTHYFIKTTTPESDLGTURLTSGRK 2591
Qy 1481 LATHILAFADGNNALIKSTSESSRIKMGVTR-----SLDREIVR 1520
Db 2592 AL-----ENGIVVYSQSTIVYNGVTRKRPADYEMQFGALALHVRGMITDEKAR 2641
Qy 1521 NERGOVTKDSRGYTDNFMKGEOALIVHGDKQFLYHTEGNKN---GK-----GPTT 1571
Db 2642 ILBQARQALABAMA-----REQQVRDGEGBARILM-TGEBKRQLLSAGKVQGYDGYV 2694
Qy 1572 RHTPE--QLVYDLKONNITDLTQ 1592
Db 2695 LSVQYQPELAD--SANNIOFLRQ 2715

RESULT 2
US-11-096-051-8
; Sequence 8, Application US/11096051
; Publication No. US20050244868A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Rameesh
; APPLICANT: MacLachlan, Timothy K
; APPLICANT: Rastelli, Luca
; APPLICANT: Vernet, Corinne
; APPLICANT: Ettenberg, Seth
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use
; FILE REFERENCE: Attorney Docket No. Cura 967
; CURRENT APPLICATION NUMBER: US/11/096,051
; CURRENT FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: 10/038,854
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 10/455,772
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: 60/557,978
; PRIOR FILING DATE: 2004-03-30
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Curaseqlist version 0.1
; SEQ ID NO 8
; LENGTH: 2725
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-096-051-8

Query Match 3.2%; Score 285; DB 7; Length 2725;
Best Local Similarity 19.5%; Pred. No. 4.4e-10;
Matches 367; Conservative 253; Mismatches 637; Indels 626; Gaps 97;
Qy 18 NEFTQANNFTSAVSGVDPRFTGLYNIQITLGHIVGNGNLQPTLPLTLSPYLNKTDIGF 77

```



Db 1155 NQFISQPPVWSIMG--NGRRRISCPSCNGQADGKLAFLVA-----LAC 1199  
 Qy 78 GI-----GPNGLSYDRKN--SLSLS-----TGNVYIE 107  
 Db 1200 GIDGSLVGDJENYVRI:FPGSNVTSLVLEKNKDFRHSNPAHRYLATDPTGDLX--VS 1257  
 Qy 108 TDKTVKQKKDLNLRFEKDLKENCYRIIHSKGDIEVLTEGPNNA--FD-----LKV 157  
 Db 1258 DMTTRIRYRK--SLTGAKDLTKNA-----EVVAGTEGCLPFDEARCGDGKAV 1305  
 Qy 158 PKKLLNPAHAIIYDWNFEATQPLNRIYDDLGDHIDPLNLEYQGLIKTILTFPGQKE 217  
 Db 1306 EATLMSPKMAV-----DKNGLIYFVDS--TWIRKVDNGIISTLL-----GSND 1348  
 Qy 218 GYRTLEAFLRQNLNSINPSLGNENPLTWSFGYPIPIGKNGILGQWTSMAFGSLKETT 277  
 Db 1349 -----LTSARPLTCD--TSMHISQVLEWPTDL--AINPMDSIY 1384  
 Qy 278 YSNNOGHHPQSANLPLVYVTLMKQVPG--AGPAIOAEXSYSHNYVGGSGSGIM 333  
 Db 1385 VLDRNVVLQITENQVRIAGRPWHCOVPGVEYVGGAAVOT--TLESATIAVSYSGV- 1441  
 Qy 334 NNKLDNLGYLMTENYSGTESRRYKDEGHQIVIRIRTYNNYHLLTSECKOON----- 387  
 Db 1442 -----LY--ITETDEKKIRIRQVTTDGEISLVA-----GIPSECDCKNDAMCDC 1484  
 Qy 388 -----GYIQOTE-----TAYVAIGH-----NF-----D 406  
 Db 1485 YQSGDGAKDAKLSAPESLASPDGTLIADLGIRIRAVSKNPLLSNMFEVASPTD 1544  
 Qy 407 SQSPQPLPKTKTETWMSADNSY-----RSBITTTEDESGNPLTKVIKD--KTKQK 456  
 Db 1545 QELYIFINGHOYVSLVGTGYLNFYSNDNDITAVT--DSNGNTL--RIRDRNRNRPVR 1602  
 Qy 457 IISPTH--WEYYPAGEVNDNCPPEYGFTRFYKKIITQTPDSEHKODPEKFIQYRSLI 514  
 Db 1603 VVEPBDNGVILTLTGNGCLK-----GMTAQGLVLFYTH----- 1637  
 Qy 515 GSGSHVTLKIBERHSATQLINSTLFOYNTDKSELGRLKQTECT-----KGENKTYIS 568  
 Db 1638 GNSGLLTKSDETEW-----TTFPYDSE--GRLVWTFPTGVVTLHAGMDXAIT 1686  
 Qy 569 VAKFTYTKDDTLQOHSITTHDNF--TIHRSQVRSYTGRLPBDTQKIVQMSYDL 627  
 Db 1687 VDIESSREEDVSI--TSNLSIDSFYTWQDQLRNSY-----QIGYDGS 1729  
 Qy 628 GRLLTRINSCTPYANTLT--YDELANLQODNRPPV--ITTDVNGQLRBEFDGAG 682  
 Db 1730 LRIT-----YASGLDSHYQTEPHVLAGTANPYAKKNMTLPGENGQNL----- 1772  
 Qy 683 RHVSQCLKSDGDKFYTIHQYDEQGRHHTSYSDYLINGRQOTDPDKVHLSMSKSYD 742  
 Db 1773 --VEMRFRKQAOQK--VNVFGRKLRVNGRNLSDVPRRTTKEKIIDHDKFL--LR:AYD 1828  
 Qy 743 NMGOIANTHWSYGVSEKITYDPIITLTATKQUSNNVQSGKEVTTTTPSQOPIQITLFD 802  
 Db 1829 TSGH--PTLWL--PSSKLMAVNVYSSTQIAS--IQRG-----TTSBK--VDYD 1870  
 Qy 803 BAGHJOSCHTLTRDG--MDRVKKTDAI-----GQCTIYQDYNRVYQITLFD----- 849  
 Db 1871 GQGRIVS--RVFADGKWSYTYLEKSNVLLHSRQYIFEDMDRSLAITMPEVAHHTM 1928  
 Qy 850 -----GTIVNRKXAPSTDTLTIDIRVNGISLQOQTFDGLSRLLTQSGDGRVMAVYTYSA 903  
 Db 1929 QTRISGYNNIYVPPESNASIITDYNEBGLL--QTAFLGTSR-----RV----- 1972  
 Qy 904 GNDQCBSTVITPQCFIHQYQPELDAVAVOASNETIQOSNYPVYTGALLKAV--AEG 960  
 Db 1973 -----LFKYRQOTLS--EILDSTRVSTFYDTAG-VLKTNLQSDG 2012  
 Qy 961 QSLPIYPSGRLKMNENIDMKKMSYLMTLGLENGYTDLGTIQKISPDTHGAVTOIKD 1020

Db 2013 FICTIRYQIGPV-----IDRQIFRFSEDEGMVNAFD-----YSYDNSFRVYSMOG 2058  
 Qy 1021 SSIKTLL-----NYDDLNRHIGS-----QVTLATGMLTTFTEPD--GLANEIGR 1064  
 Db 2059 VINETPLPDLVOPDDISGKVEQFGKGVITYDINOIISTAVWYTKHFAHRIKIEIQY 2118  
 Qy 1065 KLCDSGHTLIDIOQSWLKTQOLANRIVKLVNGVLOQRTQOYSYDSRNLNOYKCGABCEPTD 1124  
 Db 2119 EIRSLMWTITIQYD--NMGRVYKREKI--GPEANTTKAV-----EVDVQ--QLOQT- 2166  
 Qy 1125 KYGHSIVTONFYTDYIGNITACHTTPADGTEDHATEKFNANPDPCCQTEVHHTHPMDPN 1184  
 Db 2167 VYLANEKIMRYNDLNGNL-----HLNPSNSARLTP----- 2198  
 Qy 1185 IRLKYDKAGVINITD-----NHGNTENFTYDILGRQ-----NGQSGVGYD 1227  
 Db 2199 --LRYDRLDRITBLGDVQYRLDEDEGLRQGTIEFESSGGLLTRYYSKSGWTVIYRYD 2256  
 Qy 1228 PLNRIVSOQT-----DTLDCELYY--RETMVNEVANGE 1259  
 Db 2257 GLGRVYSKTSLSQHLQFPADLTYPRIITHYNNHSSSETLSLYDLQHLFAMEISSGD 2316  
 Qy 1260 MIBLRTGETIIAQORASKVLLTGTDSQOSVILLTSBKONLSQERYSAYGK--HKSTANDA 1317  
 Db 2317 -----EFYIASDN-----TGT--PLAVSSNGMLKQIQTAYGEIYFDSNIDFQ 2359  
 Qy 1318 SIIGYNGERADPVSGYHLGNGYRSYDPTLMRPHTPD-----SLSPFG----- 1360  
 Db 2360 LVIGFHGGLVDPYTKLHFGE--RDYDILAGRWTTPDIEIKRIGKDPAPFNLVYFRNNN 2417  
 Qy 1361 -----AGGINPYSYCLADPINSR--PSGHL--SWQAWTGIGMI 1396  
 Db 2418 PASKINDVQYITDVANSMLVTPGFHLNALPGFPVPEFDLTESBYELVKSQODDI--PPI 2476  
 Qy 1397 AGLLLTANGGMAIAAGGIAAIAAS-----TSTTALAEGALSVTSD--I 1439  
 Db 2477 FGVOQVYARQAKAFLSIGKNAEYQVSRBRAGAQSMLFATVXSLIGKVTALVWSQCRVQ 2536  
 Qy 1440 TSIVSGLEBPASPKASIIIG-----VWSMGMAAGLAEBSIKGTX 1480  
 Db 2537 TYNLANTANECI--KVAAVLNNAFYLENLHPTIGKQDTHYPIKTTTPSDLGTYLRLTSGRK 2595  
 Qy 1481 LATHLAFADGEGNALKTSSESRIKWYTR-----SLDREIVR 1520  
 Db 2596 AL-----ENGINVTYSOSTVYVNGTRPRADVEMQFGALALHVRKMTLDEKAR 2645  
 Qy 1521 NEBGOVICKHSRGYTDNFMKGEOAILVHGDKDFLYHTEGNRN--GK-----GPTY 1571  
 Db 2646 ILBOARQALARAWA-----REQQVRDEGEARLW--TEGKRQLLSAGKVOGYDGYV 2698  
 Qy 1572 RHTPE--QLYDYLKONNITVDLTQ 1592  
 Db 2699 LSVQYPELAD--SANNIOFLRQ 2719  
 RESULT 3  
 US-11-113-424-51  
 ; Sequence 51, Application US/11113424  
 ; Publication No. US20050260713A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gangoli, et al.  
 ; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
 ; FILE REFERENCE: 21402-225  
 ; CURRENT APPLICATION NUMBER: US/11/113,424  
 ; PRIOR FILING DATE: 2005-04-21  
 ; PRIOR APPLICATION NUMBER: 60/256,704  
 ; PRIOR FILING DATE: 2000-12-19  
 ; PRIOR APPLICATION NUMBER: 60/311,590  
 ; PRIOR FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: 60/257,314  
 ; PRIOR FILING DATE: 2000-12-20  
 ; PRIOR APPLICATION NUMBER: 60/311,613  
 ; PRIOR FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: 60/315,617  
PRIOR FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/307,506  
PRIOR FILING DATE: 2001-07-24  
PRIOR APPLICATION NUMBER: 60/322,358  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: 60/294,075  
PRIOR FILING DATE: 2001-05-29  
PRIOR APPLICATION NUMBER: 60/288,153  
PRIOR FILING DATE: 2001-05-02  
NUMBER OF SEQ ID NOS: 190  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 51  
LENGTH: 2715  
TYPE: PRT  
ORGANISM: Mus musculus  
US-11-113-424-51

Query Match 3.2%; Score 284.5; DB 7; Length 2715;  
Best Local Similarity 19.6%; Pred. No. 4.7e-10;  
Matches 302; Conservative 214; Mismatches 524; Indels 497; Gaps 80;

18 NEPTQANNFSAVSGVDPRFGLNIOITLGHVNGNLGPTLPILTSPLANKDIGF 77  
1145 NQFISQPPVSVSSIMG--NGRRRSISCPSCGQADGNKLAIPA- LAC 1189  
78 GI-----GFNPGASVYDRKN--SLSLSS-----TGENYKATB 107  
1190 GIDSLVGDVFNRYRRIFFPSGNVTSVLELNKDPHRSNPAHRYLATDPVTDGLY--VS 1247  
108 TDKTVKLOQKDLNLFPEKDLKENCYRIHKSGLIEVLTFGNNA--FD-----LKV 157  
1248 DTRNRIYRKP--SLTGAKDLTKNA-----EVAVGTGECQLPDEARCGDGKAV 1295  
158 PKKLNPAGHAIIYDWNFBATQPRLNRIYDDLDGHIDPLNLEBYOGLIKITLTFPGQKE 217  
1296 EATLMSFGMAI-----DKNGLIYFVDG--TWIRKYDONGIISTLL-----GSND 1338  
218 GYRTFLRLNQLNSIHNFSLGNENPLWSPGYTPIGKGLIGOMITSMAPAGLKEIVN 277  
1339 -----LTSARPLTCD--TSMHISQVRLWEPIDL--AINPMDSIY 1374  
278 YSNNNGHFFPOSANLPVLPVTLMKQVPG---AGOPAIQAEYSYSHNYVGGGSGNGIC 333  
1375 VLDNNVVLQITENQVRLAAGRPMHCQVGEVYPRGKAQVQT--TLBSATAIAYSGV- 1431  
334 NNKLDNLGLMTENVYSGTESRRYKDKGHDQIYRIETNNYHLLTSECQON----- 387  
1432 -----LY--ITETDEKKINRIQVLTDTGEISLVA-----GIPSECCCKDANCD 1474  
388 -----GYIOTE-----TAYVAILGH-----NF-----D 406  
1475 YQSGDGYAKADKLANPSSLAASPDGTYIADIGNIRIAVSKNKLNSMNFYEVASPTD 1534  
407 SQPSQFOLPKTKTETWRSADNSY-----RSEITETTFDESGNPLTKVIKD--KKTQK 456  
1535 QELYFIDINGHQYVSLVLTGDIYLNFSYNDNDVTAVT--DSNGWTL--RIRDDPRMVR 1592  
457 IISPSHT--WEYYPAGEVDCNCPPEPYGTRPVKKIIOGTPDSEKDKPEKFIQRYSLI 514  
1593 VVSPDNQVIMTLTGNGCLKSMTAGL-----SLVLFTH-- 1627  
515 GSQSHVTLKIEBRHYSATQNLNSTFYQYNTDKSELGRLIKOTECT-----KGENGKTVS 568  
1628 GNSGLATKSDSETG-----TTFFDIDSE-----GRLLNTYFPFGVYTNLHGMDKAIT 1676  
569 VVHKETTYTKODDTLOQSHSITTHNF--TIHSQVRSRYTGLFSDTDKDIYQMSYDKL 627  
1677 VDIBSSSEEDVSI--TSNLSSIDSEFYTMVQDLNLSY-----QGYDGS 1719  
628 GRLTLRTIANSCTPYANLTL--YDIELANLQDNRPPV---ITTDVANGQLRAEFDDAG 682  
1720 LRF-----VASGLDSHYQTEPHVLAGTANPVAKRMTLTPGENGONL----- 1762

683 RHVSQCLKDSGDGEKFTYITHTQYDEKGRHHTSTYSYDLTNGRQOTDPDKVHLSMSKSYD 742  
1763 --VEMRFRKEQAGQK--VNVFGRKLAVNGRLLLSVDFPRTTKTEKLYDHRKFL--LRALYD 1818  
743 NWQIANTHWSYVSEKITYDPIITLTKAQLOSNNSNVQTKGAVTTYTPSQOPIQITLFD 802  
1819 TSGH--PTLWL--PSSKLMAVNTYSSTQIAS-----IQRG-----TTSEK-----VDYD 1860  
803 EAGHLQSCHTLTMDG--WDRVRKETDAI-----GQCTIYQDYNRYIQTLLPD----- 849  
1861 SQGRIVS--RVPADGKTWSTTYLEKSMVLLHROQYITFEYDMMDRLSALITMSVANRHTM 1918  
850 -----GTIVNRKXAPESDTLITLDIRVNGISLGOQTFDGLSRLOSQDGRVWATYSSA 903  
1919 QTRISIGYVNNINNPESNASIITDVYBEGLL--QTAFLGTSR-----RV----- 1962  
904 GNDQCESTVTPDQGFHYQYQPELDAVLQVANSNETQOFSNVPVTKALLAV--AEG 960  
1963 -----LEKYSRRQTRLS-----EILYDSTRVSFTYDETAG--VLKTVNLQSDG 2002  
961 QSLTPIYPSGRUKMENINDMKMSYLMTLRGLENGVTDLTGITQKISRDPHGRVQIKD 1020  
2003 PCTIRYRQIGPL-----IDQIRFSEDEGMVNAFP-----YSYNSFRVTSMQG 2048  
1021 SSIKTTL-----NYDDLNRHIGS-----QVTDLATGMLTLYVEFD--GLAREIGR 1064  
2049 VINETPLPIDLYQFDDISGRVEQGRGVIIYDINOIISTAVMTYTGHPAHGRIKELQY 2108  
1065 KLCDSSGHTLIDIQOSMLTKOQLANRYKLVNGVLOFTEQYISDSRNRLNOKYKCGAECPTD 1124  
2109 EIFRSLMYWTITQYD--NMGRVYKREIKI--GPANITIKYAV-----EYDVG--QLOT- 2156  
1125 KYGHSIVTONFTYDIGNITRACHTTFPADGTEDHATFEKAPTPDCOLTEVHTHPDMPDN 1184  
2157 VYLAKEKIMRYNDLNGNL-----HLNPPSSARLTP----- 2188  
1185 IRLTYKAGRVINITD-----NHGNTENFTYDLGRLO-----NGQGSVYGYD 1227  
2189 --LRYDLRDRITRLGDVQYRLDEDEGFLRGQTEITFEYSKGLLRYVYSKSGWTVIYRYD 2246  
1228 PLNRVYQKT-----DTLDELYY--RETMUNVEVRNGE 1259  
2247 GLGRVVSSTKSLSGOHLQFPYADLTYPRTITHVYHSSSEITSLYDQHLFAMEISSGD 2306  
1260 MIRLRTGETTIAQGRASKVLTLGTDSQGSVILTSDQONLSQEAAYSANCK--HKSTANDA 1317  
2307 -----EFYIASDN-----TGT---PLAVSSNGMLMKQOTYTAGELIYPSNDVFO 2349  
1318 SILGNGERADPVSGVTHLNGVRSYDPTLMRFTHPD 1354  
2350 LVIGFHGGLVDPLTKLHFGE--RDYDILAGRWTPPD 2384

RESULT 4  
US-11-096-051-4  
Sequence 4, Application US/11096051  
Publication No. US20050244868A1  
GENERAL INFORMATION:  
APPLICANT: Kekuda, Ramesh  
APPLICANT: MacLachlan, Timothy K  
APPLICANT: Raetelli, Luca  
APPLICANT: Verneet, Corine  
TITLE OR INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use  
FILE REFERENCE: Attorney Docket No. Cura 967  
CURRENT FILING DATE: 2005-03-30  
PRIOR APPLICATION NUMBER: 10/038,854  
PRIOR FILING DATE: 2001-12-31  
PRIOR APPLICATION NUMBER: 10/455,772  
PRIOR FILING DATE: 2003-06-04  
PRIOR APPLICATION NUMBER: 60/557,978

```

/ PRIOR FILING DATE: 2004-03-30
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Cirsaeqlist version 0.1.1
/ SEQ ID NO 4
/ LENGTH: 2376
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-11-056-051-4

```

Query Match 3.2%; Score 283; DB 7; length 2376;  
 Best Local Similarity 19.4%; Pred. NO. 4.8e-10;  
 Matches 365; Conservative 225; Mismatches 637; Indels 626; Gaps 977

QY	18	NEPFOANNPTSAVSGVDPRPTGYNIGITLTHIVGNONLPTLPLSYSLNKTIDGF	77
Db	806	NOFISOQEPVYSSIMG--NGRRRSISCSGNGADGNKLAPVA-----LAC	850
QY	78	GI-----GENFGLSVYDRKN--SLSLSL-----TGENVYKIE	107
Db	851	GIDGSLVGVDDFNRYARRIPPSGNVTSVLELRNKPDRFHSNPARHYLATDVPYTGILY--VS	908
QY	108	TDKTVYLOQOKLUDLRFKDLKENQYRIHKSGLIEVLTFGRNNA--PD-----LKV	157
Db	909	DTNTRIRIYRP--SLTGAKDLTKNA-----EVAAGGEOCLPDEARCGDGKAV	956
QY	158	PKLLINPAGHAIYIDMNEATOPRLNRIYDDLDGDHIDPLMLEYOGILTKITLTPGQKE	217
Db	957	EATLMSRPMGAV-----DNGGLIYFPG--TWIRKVDNGIISTLL-----GSD	999
QY	218	GYTELRFPLNQLNSIHNFSLGENENPLTWSPGYTPIGNGLIGOMITSMTAPGKLETVN	277
Db	1000	-----LTSARPLTCD--TSMHISQVLEWPTPL-AINPMDSNIY	1035
QY	278	YSNNQGHHPQSANLPVLPVYTLMKQVPG---AGOPALAEVSYTHNVYGGSGNGIM	333
Db	1036	YLDNNVVLQITENROVRIAAGRPNGCQVPGVEYRPGKAAVOT--TLESATAIANSYSGV-	1099
QY	334	NNKLDNIYGLMTEYNGSTESRRYKDKGHQIVRIERTYNNYHLLTSECKQON-----	387
Db	1093	-----LY--ITTEDEKKIRIRIQTWTGDEISLVA-----GIPSECKDNDANGC	1135
QY	388	-----GYIQTE-----TAYYAIIGH-----NF-----D	406
Db	1136	YQSGDGAKOKKLSPSSLAASPDGTLIADIGNIRIRAVSKNPKLNSMNPVEVASPTD	1199
QY	407	QSQPOFOLPKTKETETWRGADNSY-----RSBITETTFDSGNPLTKVIK--KKTQK	456
Db	1196	QELVIFDINGHOQTVSLVGTGDYLNPSYSNDNDITATV--DSNGTL--RIIRDPRMRYR	1255
QY	457	IIPGPTH--WEYYPAGEVNCPEPPYGPTRFVKKILOTPYDSEKDPDEKFIQYRYSLT	514
Db	1254	VVSPDNQVITWLTITNGCLKSMTAQGL-----ELVLYTYH--	1288
QY	515	GSQSHVTLKIERHNSATOLNSTLOYNTDKSELGRLLKOTECT-----KGEKGKYS	568
Db	1289	GNSGLAKRSBETGH-----TTFPYDSE--GRLLNTVFPFGVYTNLHGMDKAIT	1333
QY	569	VVAKFYTKODTLQOSSHITTHDNF--TIHRSQVRSRYTGRLESPTDKDITVQMSYDKL	627
Db	1338	VDIESSSRHEEYVSI--TSLNLSISDSFYTMVQDLRNSY-----QIGYDGS	1380
QY	628	GRLLTRTLNSTPYANTPLT--YDVEYLNALQODNRPPFV--ITTDVANGNOLARNFDDAG	682
Db	1381	LRIIT-----YASGLDSHQTEPHVLAGTANFTVAKRNNTLPEGENONL-----	1423
QY	683	RHVSOCLDSDGDKFYTIHTQYVDEGRHHTSTYSDYLINGRQOTDPRKYLMSKSYD	742
Db	1424	--VEMRFKEDQAQGR-VAVPGRKALVGRNLLSVDFRPTTKTEKILYDHRKFL--LRIAYD	1475
QY	743	NMGQIANTHNSGVSEKITTVDPIITLAKQLQOSNNNTQGTGEVTTYTPSQOPIQITLFD	802
Db	1480	TSGH--PTLWL--PSSKLMANVVTYSSGQIALS-----IQRG-----TTSEK-----VDY	1521

QY	803	LAGHLOSCHTLTRGG--MDRVKXETDAI-----GQCTIYOYDNNRYIOJTLPP-----	849
Db	1522	GQGRIVS--RVFADGKTWSTYILEKSNVLLHSRQYIIFEDYMDRLSALIMBSVAHNTM	1579
QY	850	-----GTIVRRKXAPSTDTLTLDIRVNGISLQOQTFDGLSRLOSODGGRVAAVY7SA	903
Db	1580	QTRISIGYRNYINVPESMSAIIIDVNBEGHLL--QTAFLGTSR-----RV-----	1623
QY	904	GNDQCESTVITPDGQFIHQYQPELDAVALOVASNETQOQPSYVPTGALLKAV---AEG	960
Db	1624	-----LFKRRQRTLS---EILDSTRVSFTYDETAG-VLKTYNLOSOG	1663
QY	961	QSLPTPIYPSGRLXEMINIMDKKXSVMTLRLGELNGTDTLGTIOKLSRDPHGRVQIOKD	1020
Db	1664	FICTIRRQIGPL-----IDKRIFRFSEBGMVNAFD-----XYDNSFRVTSMOG	1709
QY	1021	SSIKITL-----NYDDLNRHIGS-----QVTLATAGHMLTTTVEFD--GLNREIGR	1064
Db	1710	VINETPLPIDYQDDDISGKVEGKFGVAYYDINOJLISPAVMTYXGFDAHGRKEIOY	1769
QY	1065	KLCSSSGHTLIDQOSWLKTQOLNARIYKLVGVLORTQOYSYDSSNRNLNOYKCDGAECPD	1124
Db	1770	EIPFSLWYWTIYOYD--NMGKVTREAKI--GPFANTTKYAV-----EYDVD--QLOT-	1817
QY	1125	KYGRSIYQNTFYDIYIGNITACHTTPADGTEDHATPFKPAFTDPCQLTVEHHTPMDPN	1184
Db	1818	VYLNEMKIMRKNYDNLNGNL-----HLNPSNARLTP-----	1849
QY	1185	IRLKYRAGRIVNITD-----NHGNTENFTYDLGRLO-----NGQSVYGYD	1227
Db	1850	--LAYDRLDRITRLGVDQYRLDEBGFRLQRTGEIFEYSSKGLLTRYVSKSGSGMTVIYRD	1907
QY	1228	PLANLVSGKT-----DTLCELYY--REIMLVNEVNGE	1255
Db	1908	GLGRVRSKTSYLGQHLQFPYADLTYPTRITHVYHSSSEITSLYDLOGHFLAMEISSGD	1967
QY	1260	MIRLLRTGETIIAQORASKVLLTGTDSQOVSILTDSKONLSQEAYSAYGK--HKSTANDA	1317
Db	1968	-----EPIYASDN-----TGT---PLAVFSNGLMLKQIQYTAHGEIYPPSNIDPO	2010
QY	1318	SILGNGERADPVSGVTHLNGYRSYDPTLMRFTPD-----SLSPFG-----	1360
Db	2011	LVIFGHGGLYDPLTKLHIFGR--RDYDILAGRMTTPDIEIKRICKGPAFNLVYMRANN	2068
QY	1361	-----AGINPYSYCLGDPINSD---BSGH--SWQAMVIGMG	1396
Db	2069	PASKIHVKDYITDVNSMLVTFGHLLHNAIPGPVPKFDLTBSEYELVKSQOWDI--PPI	2127
QY	1397	AGLLTLTATGMAITAAAGIYAAATAS-----TSTALAFGALSTVSD---I	1439
Db	2128	FGVOQOVAROKAKFLSLGKMAEYOVSRRRAGASWLMFAVKSJLIGGVWLVAYSQGRVQ	2187
QY	1440	TSIVSALGEDASPRASSILG-----WVSMGGAAGLAEAIKGGTK	1480
Db	2188	TNVANINAMEDCI--KYAAVLNNAFYLEMLHFTIEBKDTHYFIKTTTPBSIDGTLRLTSGRK	2246
QY	1481	LATHLGFAPBEGENALLKSTSESSERIKMGYVR-----SIDREIVR	1520
Db	2247	AL-----ENGIVTYSQSTTVVNGKRRRPRADYEMOGFALALHVRYGMTIDEBEAKR	2296
QY	1521	NBEGQVTKDHSRGYTDNFMFGKEQOAILVHGDKDFLYHTBENKJIN-----GK-----	GPYT
Db	2297	ILBQARGRALLARAWA-----REGQRRVRDDEGARLW--TBESEKQJLLSAGVQGYDGYV	2349
QY	1572	RHTBE--QVYDYLKONNIVDLTO	1552
Db	2350	LSVEQYBELAD--SANNIOFLRQ	2370

RESULT 5  
US-11-096-051-2  
; Sequence 2, Application US/11096051  
; Publication No. US20050244868A1

GENERAL INFORMATION:  
APPLICANT: Keshava, Ramesh  
APPLICANT: MacLachlan, Timothy K  
APPLICANT: MacLachlan, Luca  
APPLICANT: Verneer, Corine  
APPLICANT: Eckenberg, Seth  
TITLE OF INVENTION: Ten-43 Polypeptides and Polynucleotides and their Methods of Use  
FILE REFERENCE: Attorney Docket No. Cura 967  
CURRENT APPLICATION NUMBER: US/11/096,051  
CURRENT FILING DATE: 2005-03-30  
PRIOR APPLICATION NUMBER: 10/038,854  
PRIOR FILING DATE: 2001-12-31  
PRIOR APPLICATION NUMBER: 10/455,772  
PRIOR FILING DATE: 2003-06-04  
PRIOR APPLICATION NUMBER: 60/557,978  
PRIOR FILING DATE: 2004-03-30  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: CuroSeqList version 0.1  
SEQ ID NO 2  
LENGTH: 2715  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-11-096-051-2

Query Match 3.2%; Score 283; DB 7; Length 2715;  
Best Local Similarity 19.4%; Pred. No. 5.9e-10;  
Matches 365; Conservative 255; Mismatches 637; Indels 626; Gaps 97;

18 NEFTQANNFTSAVSGVDPRTGLYNIQITLGHIVGNGMLPPLTSLSPKMTDIF 77  
1145 NQFISQCPVVSIMG-NGRRRSISCPSCGQADGNKLAPVA-----LAC 1189  
78 GI-----GNFGLSVYDRKN--SLSL-----TGENYKIE 107  
1190 GIDGSLVYGDNYVARIPTSGSVNVSLELRNKDFHSSNPAHRYLATDPTVGLDLY--VS 1247  
108 TDKTYKLOOKLMDLRFKEDLKENCYRIHKSGLIEVLTFGNNA--FD-----LKV 157  
1248 DYNRRIRYRK--SLTGAKDLTKNA-----EVAAGTGEQCLPDEARCGDGGAV 1295  
158 PKKLINPAHAIYIDWNEFATQPRINRIYDLDGHDIFLNLERYOGLIKITLTPPGKE 217  
1296 EATLMSPKGMAY-----DNGLIYFVDG--TWIRKYDONGIISTLL-----GSND 1338  
218 GYRTLERLANQNLNINFSIGENENPLTMSGYTPGIGKGLIGQITSMTPAGLKEIVN 277  
1339 -----LTSARPLTCD--TSMHISOVRLMPDCL--LNPMDNIT 1374  
278 YSNNQGHHPPOSANLPVLPVTLMLKQVPG--AGPAIOAESYTSNHYVGGSGNGIM 333  
1375 VLDNNVVLQITENRQVRIAAGRPMHCQVPGVEYVPGKHAQOT--TLESATLAVSYSGV- 1431  
334 NKKLDNLGLMTEVYSGTESRRYKDGHDQIVRIERTYNNHLLTSECKQON----- 387  
1432 -----LY--ITETDEKINRIQVTTDGEISLVA-----GIPSECCCKDANDC 1474  
388 -----GYQTE-----TAYYAIIGH-----NF-----D 406  
1475 YOSGDGVAKDALSPSSLLASPDGTLIADIGNRIRAVSKKPKLNSMFEYVASPTD 1534  
407 SOPSOFLPKTKETETWRASDANSY-----RSEITETTFDESQNLTKVIKD--KTKQK 456  
1535 QELYFDINGHTQYVSLVTDGYLNPFSYNDNDITAVT--DSNGTL--RIRDRPMBVR 1592  
457 IISPESTH--WEYYPAGEVNDCPREPYPGFTFRVKKIOTPYDSEKDPKFIQYRSLI 514  
1593 VVSPDNQVITLITGNGCLKSWTAQGL-----ELVLFTHY-- 1627  
515 GSGSHVTLKIERHRSATQOLNSTLFOYNTDSEHGLLKQTECT-----KGENGKYVS 568  
1628 GNSGLLARKSDETGW-----TTFDYDS--GRLLNVPFPGVVNLHGDMDKAIT 1676  
569 VNHKFTYTKODDTLQOSSHITTHDNF--TIHRSQVRSRYTGRLLFSDTDKDIYQMSYDKL 627

1677 VDIESSEREEDVSI--TSNLSSIDSFYTWQDDLRNSY-----QIGVDGS 1719  
628 GRLLTRTLNCGTYRANTLT--YDELANLODDNRPPV-----ITTDVNGQLRNRPFGAG 682  
1720 LRIT-----YASGLDSHYQTEPPVLAGTANPVAKKNMMLPGENGNL----- 1762  
683 RHYSQCLKSDGQCKFTHTHQYDEQGRHNTSTYSPLTNGRQQTDPDKVHLSMSKYD 742  
1763 --VEMRRKQQAQK--VNVGRKLAVNGRRLSVDPRTTKTEKIYDHRKFL--LRITAD 1818  
743 NMQIANTHWSYGVSEKITVDPTLATKQLQSSNNVQGEKVTYTPSQOPIQITLPD 802  
1819 TSGH--FTLWL--PSSKLMAVNVYSSTQIAS-----IQRG-----TSEK-----VDYD 1860  
803 EAGHLQSCHTLTNDG--WDRVKETDAI-----QOCITQYDNNRYQITLPD----- 849  
1861 GQGRIVS--RVFADGKTWSYTYLEKSNVLLHSORQYIFEDWMDRLSATIMPSVANRHTM 1918  
850 -----GTIVNRKAPSTDTLITDIDKNGISLQOQTFDGLSRLTQSDGGRVAAVYISA 903  
1919 QTRISGYRNIYNPSSNASITTDVNEGLL--QTFLETSR-----RV----- 1962  
904 GNDQCPSTVITPQGQFIHQYQPELDAVLQVANSNEITQGFSPNPTGALLKAV--AEG 960  
1963 -----LFTYRRQTRLS-----EILYDSTRVFTYDETAG--VLKTVNLQSDG 2002  
961 QSLPIYTPSGRLKMEINIMKMSYMTLRLGLENGTDLTGITQKISRDRHGVTQIKD 1020  
2003 FICTIRRAQIGPL-----IDROIIFRSBDGMVNAF--YSSNPSFTVSMQG 2048  
1021 SSIKTL-----NYDNLNRHGS-----QVTDATGCMULTTYEFD--GINREIGR 1064  
2049 VINETPLPIDLQCPDLSGKVEQGRGVITYINDQISRAWMTYTHGFPAHGRIKIYQ 2108  
1065 KLCDSGHTLIDIQOSWLKTOQLANRIVKNGVLQRTFOYSYDSNRNLNQYKCDGAECPD 1124  
2109 EIFRSLMWITLIQVD--NMGRVTKREIKI--GPEANTTKYAV-----EYVDG--QLQT-- 2156  
1125 KYGHSIVTONFYDNIYGNITACHTTFADGTEADATFPFANPTDCOLTEVNHTRPMDPN 1184  
2157 VYLNKIKMWRKYNVDLNNL-----HLNPSNARLTP----- 2188  
1185 IRLKYDKRAGVINTD-----NHGNTENFTYDILGRLO-----NGQGSVGYD 1227  
2189 --LKYDLRDKITRIGDVQYVLDEDEGFLRQGTETEFESKGLLRVYSKSGMTVIIRYD 2246  
1228 PLNRLVSQKT-----DTLDCELY--RETMLVNEVRNGE 1259  
2247 GLGRVSSKTSLQHLQFYADLTYPTRITHVYHSSSEITSLYDYDQHLFAMEISGD 2306  
1260 MIRLRTGETTIAQORASKYLLTCTDSQGSVILTSDQONLSQEAYSVAGK--HKSTANDA 1317  
2307 -----EYFIASDN-----TGT-----PLAVFSSNGMLKQIOYTYAGELIYFSPSNDIFQ 2349  
1318 SILGNGERADPVSGVTHLNGYRSYDPTLMRFTPTD-----SLSPFG----- 1360  
2350 LVIGFHGLYDPLRLKLIHPGE--RDYDILAGRWTPPIELIKRIGKOPAPFNLMFRRNN 2467  
1361 -----AGGINPYSYCLGDPINSD--PSGHL--SWQAMTGIGMI 1396  
2408 PASKIHVKDYITDVNSMLVTFGFHLNNAIPGFVPKFDLPEPSYELVKSQOMDPI--PPI 2466  
1397 AGLLTTATGMAALAAAGTAAALAS-----TSTTALAFLGALSVSD--I 1439  
2467 FGVOQOVARQAKAPLSGKMAEYQVSRRAAGAQSMWIPAVKSLIGGVMLAVSQGRVQ 2526  
1440 TSIVSGLALEDAPRASISILG-----VWSMGAGAAGLAESAIRKGTGX 1480  
2527 TNVNTINANEDCI--KYAAVLNNAFYLENLHPTIEBKDTHYFIKTTTPSBDLGTALNTSRK 2585  
1481 LATHIGAFAEDGENALLKSTSESSRIKMGVTR-----SIDREIVR 1520

```

Db 2586 AL-----ENGINEVTVSOSTTVVNGRRRPADEVMOFGALALHVRIGMTDESEAR 2635
QY 1521 NEEGQVYKHSRGYTDNFMKGEGQALLVHGDQGFYHTEGKRN-----GK-----GPT 1571
Db 2636 ILBQARQALARAAMA-----REQQVRDEBEGARLM-TEGEEKQLLSAGKVQGYDYV 2688
QY 1572 RHTEP--QLVDYLDKNNIVDLTQ 1592
Db 2689 LSVEQYPELAD--SANNIOFLRQ 2709

RESULT 6
US-10-821-234-1097
/ Sequence 1097, Application US/10821234
/ Publication No. US2005025511A1
/ GENERAL INFORMATION:
/ APPLICANT: Labac, Ivan
/ APPLICANT: Seache-Crain, Birtglt
/ APPLICANT: Andarmant, Susan
/ APPLICANT: Tang, Y. Tom
/ TITLE OF INVENTION: Methode for Diagnosis and Treatment of Preeclampsia
/ FILE REFERENCE: 821A
/ CURRENT APPLICATION NUMBER: US/10/821,234
/ CURRENT FILING DATE: 2004-04-07
/ PRIOR APPLICATION NUMBER: US 60/462,047
/ PRIOR FILING DATE: 2003-04-07
/ NUMBER OF SEQ ID NOS: 1704
/ SOFTWARE: PL_SEQ_genes Version 1.0
/ SEQ ID NO 1097
/ LENGTH: 1094
/ TYPE: PR1
/ ORGANISM: Homo sapiens
US-10-821-234-1097

Query Match 3.1%; Score 273.5; DB 6; Length 1094;
Best Local Similarity 20.4%; Pred. No. 5,7e-10;
Matches 255; Conservative 164; Mismatches 417; Indels 413; Gaps 62;

QY 594 FTTHRSQVRSRYGRFSDTDKDIYQMSYDKGLRLTLTNSGT-----PYANTLTY 647
Db 3 FTHGN-----SGLLATKSDDEGTMTFPDYDSEGRLTNVTFPGVVNTNLHGDMDKAITV 56
QY 648 DYELNNLQDNNRPFTTTDVTG-----NQLNNE-----PDGGRHVSOCLXDS-- 692
Db 57 DISSSKREED-----VSITNSLSIDSFYTMVODQLNNSYQIGDGLRLITVASGLDSHY 111
QY 693 -----DGDGKFYTHIQOYDEQGRHHTSYSDYL-TNGR----- 725
Db 112 QTEPHVLAGTANPTVAKRNMTLPGENGONLVEWFRKEQAQKKNVGRKLRVNGRLILS 171
QY 726 -----QOTDPKVLHSMK-----SYDNWGOIANTHWSGVSEKIYVDPITLTLATKQLQSN 776
Db 172 VDEDRITKTEKIDYDHRKFLRIAYDTSGH--PTLML--PSSKLMANVTYSSTGTQIAS- 226
QY 777 SNNVQGEKVTYTPSQOPIQITLFEAGHLQSGHILTRG--WDRKKTDAI-----G 829
Db 227 ---IQR-----TTSK-----VDYDQGRIVS--RFPAQKWTSTYTLKSMVLLHSQ 271
QY 830 QCTIYQDYNRVQIQLPD-----GTIVNRKYAPFSTDTLITDIRVNGISLG 877
Db 272 ROYIFEDMDRLSAITMPSVAHRTMOTIRSIGYRRIYNPBPENASIIITDYNBEGILL- 330
QY 878 QCTPFDGLSLTQSDGGRWAAVYYSAGNDCPSTVITPDQFIHYQVQPELDAVLQVAS 937
Db 331 QTAFLGTSR-----RV-----LFKYRQTLBS-----ELLY 356
QY 938 NEITQOESSVPMYTGALIKAV---AEGSLPIYVPSGRLMENINDMKKMSYMLTJLGL 994
Db 357 DSTRVSTFYETRG-VLKTYNLQSDGFTCTIRKQIPL-----IDKQIFRSEBGMV 408
QY 995 NGYTLGLTLOKISRDTGHRVTOIKDSIKITTL-----NYDDLRAHIGS-----QV 1040
Db 409 NARD-----YSYNSFRYTSWQGVINETPLPIDLYQDIDISGKKEQGGKRGVYYDI 461

```

```

QY 1041 TDLATGMLTTYEPD--GLNREIQRKLCDSGHTLIDIQOSWLKTQQLANRYKANGVLQ 1098
Db 462 NQIISTAVMTYTGHPAHGRIKEIOYEIRFSLMYWITTIQYD--NMGVTVRREIKI--GPA 518
QY 1099 RTEQVYSDSNRNLNOVKCDGACBPDKYGHISITQNTQYIYIGNITACHTTPADGTEDHA 1158
Db 519 NTKKTAI-----ETVDG-QLOT-VYLNKIKMWRINYDLNGL----- 554
QY 1159 TPKRANPTPCQULEVHHTHPDMPDNIRLKYDARAVINITD-----NHGNTEN 1207
Db 555 --HLNPSNSARLT-----LNYDLDRITRLGDVQYRLDEDEPLRQRTETI 599
QY 1208 FTYDTLGRLO-----NGQSVYGYDPLNRYVQKT----- 1237
Db 600 FEYSSKGLTRVYKSGSGMTVIYRYDGLGRVBSKTSIGQHLQFPYADLTYPRTITHVN 659
QY 1238 --DTLDCELY--RETMVNEVANGEMIRLRLRGETIIAQGRASKVLTLGTDSQSVILT 1293
Db 660 HSSSEITSLTYDLOGHLPAMEISSGD-----EFTIASDN-----TGT--PLAVFS 702
QY 1294 SDRONTLQEAAYSAYGK--HKSTANDASILGYNERADPVSVYTHLNGVYSYDPTLMRFH 1351
Db 703 SNGMLMKQIYTAVGEIYFDSNIDFQLVIGFHGLYDPLKLIHFG--RDYDIIAGRW 760
QY 1352 TPD-----SLSPFG-----AGINPYSYCLGDP 1374
Db 761 TPDIEIKWKRIKGDPAPEPNLYMFRNNPASKIHDKVQYITDVNSWLVYFGHNLNADPFP 820
QY 1375 INRSD--PSGHL-SMQAWTIGMGAGLTLTATGSMALAAAGTAAALAS----- 1422
Db 821 VPKFDLTREYELVKSQOMDI--PPIFGVQOVARQAKALISGKMAEVQVSRRRAGAQ 879
QY 1423 -----TSTALAFGLSVTSD--ITSIVSGLBEPASSILG----- 1459
Db 880 SWMLPATVSLIKGWMLAVSQGRVQTNVIANEDCI-KVAAVINNAFYLENHLFTIEG 938
QY 1460 ----WVSMGMAAGLAESAISKGTALATHGLPAEDGENAALKSTSESRIRKGVTR-- 1512
Db 939 KOTHYPIKTYTPESDGLTURLTSGRKAL-----ENGINEVTVSOSTTVVNGRRRP 988
QY 1513 -----SLDREIVNBEQVYKHSRGYTDNFMKGEGQALLVHGDQXG 1554
Db 989 ADVEMOFGALALHVRIGMTDESEKARILEQARQALARAAMA-----REQQVRDEBEGA 1042
QY 1555 FLYHTEGKRN-----GK-----GPTRHTEP--QLVDYLDKNNIVDLTQ 1592
Db 1043 RLM-TEGEEKQLLSAGKVQGYDYVYLSVEQYPELAD--SANNIOFLRQ 1088

RESULT 7
US-10-453-372-170
/ Sequence 170, Application US/10453372
/ Publication No. US20060003323A1
/ GENERAL INFORMATION:
/ APPLICANT: Alsobrook, et al.
/ TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
/ FILE REFERENCE: 21402-589 A
/ CURRENT APPLICATION NUMBER: US/10/453,372
/ CURRENT FILING DATE: 2003-06-03
/ PRIOR APPLICATION NUMBER: 09/789390
/ PRIOR FILING DATE: 2001-02-23
/ PRIOR APPLICATION NUMBER: 60/185967
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: 09/823187
/ PRIOR FILING DATE: 2001-03-29
/ PRIOR APPLICATION NUMBER: 60/195792
/ PRIOR FILING DATE: 2000-03-10
/ PRIOR APPLICATION NUMBER: 09/839446
/ PRIOR FILING DATE: 2001-03-19
/ PRIOR APPLICATION NUMBER: 60/199476
/ PRIOR FILING DATE: 2000-03-25
/ PRIOR APPLICATION NUMBER: 09/863776

```

```

; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining prior application data removed - See file wrapper or PAM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Cureseqlet version 0.1
; SEQ ID NO 170
; LENGTH: 2333
; TYPE: PR1
; ORGANISM: Homo sapiens
; US-10-453-572-170

Query Match      3.1%; Score 272.5; DB 6; Length 2333;
Beat Local Similarity 19.5%; Pred. No. 2.3e-09;
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

QY 49 GHIYNG-NIGPTLPPLTSLSPKLTIDIGFIGNFGLSVYDRKNSLSTGEHYK--- 104
DB 666 GHLEFKSFOASPNLSTFTW---DKTDA-----YQRYVGLSDAVSV--GEYEETCP 713
QY 105 --VIETDKTVLQOKKID--NL---RPEK---DLKENCYRIHK--SGDIEVLTG----- 147
DB 714 SLIMEKRTALLQFELDPNLSGMSLDKHHILNVKSG---ILKGTGENQPLTQOPALI 770
QY 148 ---FNNAFDLKV-----KLLNPAGHAIIYIDMFEATOPRLNRIYDDLDGDHDP 195
DB 771 TSIMNGRRRSISCPSCNGLAEGNKLALPALAVGIDGSL----- 810
QY 196 LLMLEYOGLITLITLPFGCKEGRTERFLNRQL---NSIHFSIGENMLTWSFGYT 251
DB 811 ---YGDENYIRIRPEPRNVTSILELR--NKEFKSNNPAHKXYLA--VDPVSGSLYVS 862
QY 252 PIGKNGILQWITMTAPGKLETVYNSNNQHNFPQSANLPVLPYVTLAKQVPGAGOP 311
DB 863 DTNSRRIT---RVKSLSGTKDLAGNSEVVA---TGEQCLPF---DEACGPGGK 908
QY 312 AIOAEYSTSHNVYGGSGNGIWN-----NKLNLVGLMTEYVYGGTESRRYK---DK 360
DB 909 AIDA--TMSBRGIAVNDKGLMFEVDATMIRKVDQ--NGIISTL--LGSNDLTAVRPLSCDS 964
QY 361 EGHQOIVRIE-----RTYNNHLLTSECKQONGYIQTETRYVALIIGNPFSQSQF 412
DB 965 SMDVAQVLEMPDTLAVNPMDSLYVL-----ENNVIIRITENHOVSIIA---GRPMHC 1015
QY 413 QLPRTKETWRSADNSYRSEITETTFDESGNPLTFVIRKDKTKOKIISPSTMEYYPAGE 472
DB 1016 QVPGIDYSLSLALHSLAESASALASHTGVLVYITERDEKKINRLQVTTNGEICLLAGA 1075
QY 473 VDNCEPPE-----PYGFTRFVKKIITQPYDEBFKDP 503
DB 1076 ASDCDCKDVANCYSGDAAYTDAIILNPSLSLAVAPDG--TIYIADLNLIRIRAVSKNP 1134
QY 504 --EKPEIORYSLIGSOSHVLTKIBERHYSATQOLNSTL---FOVNTDSEKSGRLIKQEC 558
DB 1135 VLANAQVEASPGQELVFNADSIHQYTVSLVGEVLYNFTYSTD--NDVTELI----- 1188
QY 559 TKGENGKTSYVNHKEFTYTKODDTLQOSH-----SITTHDNFTIHSQ----- 600
DB 1189 --DNNGNSLKT-----RDSGSGMRHLLMPDNOIITLVGTNGGLKLVVSTQNLIELGM 1239
QY 601 VRSRYGRLPSPDTYDITVQMSYDKLGRLLTRITNSGT-----PYANTLTUYELANL 654
DB 1240 TYDNGTGLLATKSDTGTFTFYDYDHEGRLTNVTPGVVTSILHREMEKSIITIDIENSR 1299
QY 655 QDDNRPRPVITTTDVG-----NOLRNEFD-----GAGRH----- 684
DB 1300 DDD-----VTYITNLSSVEASVTVVQDOVRNSYQLCNNGTLRLVMYKNGKGISFSEBPVL 1354
QY 685 -----VSQC-----LKDSGDGKFTYIHTQYDEQGRHNTSTYSYDL 721

```

```

DB 1355 AGTITPTIGKCNLSLPMENGNSIEWRLRQKGR-VTFGRKLHGHNNLLSI--DYD 1411
QY 722 TNGR--QOTBPDVKHLMSKSYDNMGQ-----IANTHWSY-----GVSEK 759
DB 1412 RNIRTEKIIDDKHKKFLRIIYDVGPRFLMLPSSGLAAVNVSYFFENGRLAGLORGANISER 1471
QY 760 ITVDPITLATKQLOSNNSNVQTEKEVTTYTPSQPQITLTFPEAGHLQSCHTLTRDGM 819
DB 1472 TDID-----KQGRIVSRMFADGK--VMSGYLCKSMVLLQSGORQ----- 1510
QY 820 RVRKETDAIGQCTIYOYDYNRVLIQTLPDGTIVNRKAPFSTDTLITDIRVNGISLGGQ 879
DB 1511 -----IFEXDSDRLAAVTMPS---VAHH--SMSTHTSIGYIR-----NIYNP 1548
QY 880 TFDGLSRLOSQDGRVMAVATYSAGNDQCPSTYITTDGQRIHYQ--QPELDAVYLQVAS 937
DB 1549 PERNASVIFPDSYDDGIL-----KTSFLGTGRQVFKYKGLSLSEIYVD--S 1594
QY 938 NEITQGPSYNPVTGAL--LKAVALGOSLPRIYPSGRLKKNENINDMKMSYLTLRGLN 995
DB 1595 TAVT--FGYDETTGVLMVNLQSGCFSTIRKIGEL-----VDKQIYRSEBGMVN 1645
QY 996 GYTDLTGTLQKISRDTHGRTVQIKDSIKTTL-----NYDDLN--RHIGS-----QVT 1041
DB 1646 ARPDYT-----YHDSFRIASIKPVLSERPLPVDLRYRDBISGKVEHFGKFGVYYDIN 1699
QY 1042 DLATGHMLTTTVPED--GLNREIGRKLCDSSGHTLDOIQSWLMTQOLANRIVKLN---GV 1096
DB 1700 QITTTAVMTLSKHFDTGRIKEVOYEMFRSLMTMTVQYDSM-----GHVIRELKLGP 1753
QY 1097 LQRTQGYSDSRNRLNQKCDGA-----ECPDXYGHSIVTQNFYDIYGNITACHTT 1149
DB 1754 YANTTKKTYD-----YDSDGQLQSAVANVDRFTMR-----SYDLNGNL----- 1791
QY 1150 PADGEDHATFKRANPTDPCQLTEVHNHPDMEDNIRLKTDKAGRVNI-----TNGK 1203
DB 1792 -----HLNPNNSVRL-----MP-----LRVDLDRITRLDVOYKXIDDG 1827
QY 1204 -----NTENFTYDTLGL-----QNGQGSVGYDPLNRLVSOQTDPLDCELYYRETMV 1252
DB 1828 YLCQSGSDIFEVNSKGLTALYANKASGWSVOYRDKGRASAYKTN--LGHNLQFYSDLH 1886
QY 1253 NEVR-----NGEMIRLL--RTGETIILAQBPASKVLLGTDSQOS--VILTSDKONLS 1300
DB 1887 NPRTITHVNHNSSEIRSLVYDLOGHLFAMESSGEYVYUASDNTGPRPLAVFSINGMLIK 1946
QY 1301 QEAYSAYGK--HKSTANDASILGNGERADPVSCVTHLGNGYSYDPTLMKRFHTPD--- 1354
DB 1947 QLOQTAYGEIYYDSNPFCQWVIGFHGGLYDPLTGLVHFTQ--RDYDVLAGRWTSPDYTMW 2004
QY 1355 -----SLSPFGAGSINYSYCLGDPINRS--DPSGHL-----SMQWMTGI----- 1392
DB 2005 KNGKKEPAPF-----NLMYFRSNNPLSSELDKMYIVTDVKSMLVMPFGOLSNITPGPPRA 2059
QY 1393 -----GWSIAGLLITATGMAIAAG-----IAAALA----- 1421
DB 2060 KMYFVPPRYELSESQASENGLITGVQOTTERBHQAFLALEGQYITKKLHNSIEKAGHW 2119
QY 1422 -SISTTALAPGAL-----SYTSDITSIVGALLEDAPPKASSILG----- 1459
DB 2120 FATTPPIIGKIMAIKEGRVTTGVSSIAS---ED--SRKVASVLANNAUYLDKMYHSIEBK 2175
QY 1460 -----WVSMGMAAGLA-----ESA-----IKGSKLATIHGAFAEDGE 1493
DB 2176 DTHYFVKGISAGDGLVTLGTTIGKAYLESQVNTVSOPTLLVNGKTRAFNI---EQY 2231
QY 1494 NALIKSTSESSIRIKMGVT--RSLREIVRNEBGVYIKDHSRGYTDNFMKGQOAILVHGDK 1552
DB 2232 STLLS-----IKYGLTPTLDEKAR-----VLQARQALGTAMAKSQQKARBQ--R 2278
QY 1553 DGLYHTTEGK 1563

```

Db 2279 EGSRLMTEGEX 2289

## RESULT 8

US-10-453-372-114

Sequence 114, Application US/10453372

Publication No. US2006003323A1

GENERAL INFORMATION:

APPLICANT: Alebrook, et al.

TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

FILE REFERENCE: 21402-589 A

CURRENT APPLICATION NUMBER: US/10/453,372

CURRENT FILING DATE: 2003-06-03

PRIORITY APPLICATION NUMBER: 09/789390

PRIORITY FILING DATE: 2001-02-23

PRIORITY APPLICATION NUMBER: 60/185967

PRIORITY FILING DATE: 2000-03-01

PRIORITY APPLICATION NUMBER: 09/823187

PRIORITY FILING DATE: 2001-03-29

PRIORITY APPLICATION NUMBER: 60/195792

PRIORITY FILING DATE: 2000-03-10

PRIORITY APPLICATION NUMBER: 09/839446

PRIORITY FILING DATE: 2001-03-19

PRIORITY APPLICATION NUMBER: 60/199476

PRIORITY FILING DATE: 2000-03-25

PRIORITY APPLICATION NUMBER: 09/863776

PRIORITY FILING DATE: 2001-05-23

PRIORITY APPLICATION NUMBER: 60/208263

PRIORITY FILING DATE: 2000-05-31

PRIORITY APPLICATION NUMBER: 09/939398

PRIORITY FILING DATE: 2001-08-24

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

PRIORITY APPLICATION NUMBER: 60/227800

PRIORITY FILING DATE: 2000-08-25

QY 413 QLPKTEWRSADNSRSEITETTPESGNPLTKVTKDKTKKIIISPSTHMEYYPAGE 472

Db 1345 QVPEIDVSLKLAHSHLESASALAIASHTVLYITETDEKKNINLRQVTTNGECLLAGA 1404

QY 473 VDNCPPE-----PYGFRFVKIILQTPYDESPDDP 503

Db 1405 ASDCDKDNVNCYSGDDAVDAIDAIINSSSLAVAPDG-TIYADICNIRIRAVSKRP 1463

QY 504 -EKPIQRYSLIGSGHVTLKIEBRYASATOLLNSTL---FOYNTDKSELGRLLKOTEC 558

Db 1464 VLNAFNQYEAASPOEQLYVFNADGIHQYTVSLVGYELVNFYSTD-NVYTELI----- 1517

QY 559 TKGENGTYSVAKFTTKKODDTLQOSH-----SITTDNFTIRSO----- 600

Db 1518 -DNNGNSLKI-----RDSGGMPEHLLMPDQIITLLTVGNGKLKVVSTONLEGLM 1568

QY 601 VRSRYGRLPSPDPTKDIIVTQMSYDKLGRLLTRPLNSGT-----PYANTLTVOELNLL 654

Db 1569 TYDGNITGLTKSDDEGTWTFPYDYDEGRLLTNVTRPGVVTSLHREMEKSIITDIENSNR 1628

QY 655 QDDNRPPVITTTDVNG-----NQLRNEPD-----GAGRH----- 684

Db 1629 DDD-----VTVITNLVSVEASVTVQOVQVANSYQLCNGGTLRVVYANGMGISFHSBPHVL 1683

QY 685 -----VSOC-----LKSDGDKTYTHTQYDEQGRHHTSYGDL 721

Db 1684 AGTTPPTIGRCNLSLPENGNLSIEMLRREKQIKG-VTFGRRLVHGRLSLI--DYD 1740

QY 722 TNGR-QQTDPPKVLHSMKSVDNMWQ-----IANTMSY-----GVSEK 759

Db 1741 RNIRTEKIYDHRKFTLRIITDYQGRFPLPSSGLAANVSTYFNGRLAGLGAMSER 1800

QY 760 IYVDPITLTKQLQSNNSNVQGEVYTYTPSQPIQITLFBEGHLAGCHTLTRDGD 819

Db 1801 TDID-----KQRIYSRMFADOK-VMSYSLDKSMVLLQSQROQ----- 1839

QY 820 RVKREPDALGCTIYQVDNANRYQITLPDGTIVNRKAPSPDTLLTDIRVNGISLQO 879

Db 1840 -----IFEDSSDRLLAVTMP--VAAR--SMSTHTSISGYIR--NTYNP 1877

QY 880 TFDGLSLTQSGQGRVWATYSAGNDQCPSTVITPPGQPIHYOY--QPEIDDAVLAVAS 937

Db 1878 PESNASVIFPYSDDGRL-----KTSFLGGRQVFPYKYGSLSLSELYVD--S 1923

QY 938 NEITQGFSPNPTGAL--LKAVALGOSLPRIYPSGRLKMNENDMKMSYLTWLRGLEN 995

Db 1924 TAVT--FGYDETTGVLMKVMVLQSGGFGCTIRYRKIGPL-----VDKQYRFBEGAMVN 1974

QY 996 GYTDLCTIOKISRDTGRVTOIKDSIKTTL-----NYDDL--RHIGS-----QVT 1041

Db 1975 ARPDYT-----YHDSFRASIKPVLSEPLVDLYRYBESIKVGFPGFVLYYDIN 2028

QY 1042 DLATGMLTTTTPED--GLNRBIGRLCDSSGHTLDIQOGLMQLQOLANRIVKLN--GV 1096

Db 2029 QIITTAVMILSKFIDTGRKKEVOYEMFRLMTMTVQYQSM-----GVYIKRELKGP 2082

QY 1097 LQRTQSYDSSRRNLQYKCDGA-----ECPDKYSGHIVTQNFYDIYGNITACHTT 1149

Db 2083 YANTTKTYD-----YDGDQLQSVAVANDRPTMR-----SYDLGNML----- 2120

QY 1150 PADGTEDHATFKANPDPQOLTEVHNHTHPDMDNIRLTKDKAGRVNLT-----TNNG 1203

Db 2121 -----HLNPNNSVRL-----MP-----LRYDLDRITRLDVOYKIDDDG 2156

QY 1204 -----NTENFTYDTLGLT-----ONGQGSVYGDPLNRLVSOQTPDLDLRYRETMVL 1252

Db 2157 YLCQSGSDIFENYSKGLTRAYNKASGMSVOYRTDVGRAASIKTN--LGHNLQYFYSDLN 2215

QY 1253 NEVR-----NGEMIRLL--RTGETTIAQORASKVLLTGTDSQOS--VILTSDKONLS 1300

Db 2216 NPRTITHVNHNSSEITSLYDQGLHFLAMESSSGEEYVAVASDNTGPRPLAVFSINGIMIK 2275



QY 1301 QEAYSAYGK--HKSTANDASILGNGERADPVSGVTHLNGRSYDPTLMKFTTPD----- 1354  
Db 2276 QLOQTAYGEIYYDNDPQWVIGFPGGLYDPLTKLVHFTQ--RDYDVLAGKMTSPDITMW 2333  
QY 1355 -----SLSPFAGAGINPYSYCLGDPINRS-DPSGHL-----SWQAMTGI----- 1392  
Db 2334 KNGVEKPAFP-----NLTFKSNPNPLSELDLKNVYVTVKSWLVWFQFOLNIIIFGFPRA 2388  
QY 1393 -----GNGIAGLLITITATGMAIAAAG-----IAAIA----- 1421  
Db 2389 KMVYVPPPELYELSESQASENGQLITGVQOTTERHNOAFMALEGQVITTKGLHASIREKAGHW 2448  
QY 1422 -STSTTALAFAAL-----SVTSDITSIVSGALEDASPASISIG----- 1459  
Db 2449 FATTPITIGKIMPAIKRGRTTGVSSIAS---ED-SKVASVLMNNAVYLDKMHYSIEGK 2504  
QY 1460 ---WVSMGMAAGIA-----ESA-----IKGSKLATHLGAPAEDE 1493  
Db 2505 DTHYFVKIGSADGDLVLTGTTIGRKVLESQVNVTVSQPTLLVNGRTREFINI---EFQY 2560  
QY 1494 NALKSTSESRIRIKGVT--RSLDREIVNREGQVYIKHDSRGVTDNFMKGKGOAILVHGDK 1552  
Db 2561 STLISL-----IRYGLTPDTLDEKAR-----VLDAQROPALGTAMAKEQOARDS-R 2607  
QY 1553 DGPLYHTEGNK 1563  
Db 2608 EGSRLMTEGK 2618

RESULT 9  
US-10-453-372-148  
; Sequence 148, Application US/10453372  
; Publication No. US2006000323A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook, et al.  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
; FILE REFERENCE: 21402-589 A  
; CURRENT APPLICATION NUMBER: US/10/453,372  
; CURRENT FILING DATE: 2003-06-03  
; PRIOR APPLICATION NUMBER: 09/789390  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/185967  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 09/823187  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195792  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/839446  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/199476  
; PRIOR FILING DATE: 2000-03-25  
; PRIOR APPLICATION NUMBER: 09/863776  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: 60/208263  
; PRIOR FILING DATE: 2000-05-31  
; PRIOR APPLICATION NUMBER: 09/939398  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: 60/227800  
; PRIOR FILING DATE: 2000-08-25  
; Remaining Prior Application data removed - See file Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1609  
; SOFTWARE: CureseqList version 0.1  
; SEQ ID NO 148  
; LENGTH: 2724  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-453-372-148

Query Match 3.1%; Score 272.5; DB 6; Length 2724;  
Best Local Similarity 19.5%; Pred. No. 2.9e-09;  
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

QY 49 GHIVNG-NLGPFLTLTSLSPNLNTDIGFGIGFNGSLVVDKRNKSLSLSTGENYK--- 104

Db 1057 GHLFQKSFQASPMILASFPIW---DKTDA-----YQARVYGLSDAVAVS---GREYETCP 1104  
QY 105 --VLETDKYKLOOKKD--NL--RPEK-----DKENCRIIHK--SGDIEVLVG----- 147  
Db 1105 SLIEMKRYTALLOQFELDPNSLQWSLDKHIILNVSG---ILHKGTGENQOFLQOQPAII 1161  
QY 148 -----FNNNAFDLKP-----KGLNPAGHAIIYDMNEFATQPRINRYDLDLGDHDP 195  
Db 1162 TSTMGNGRRRSISCPSCNGLAEGNKILAPVALAVGIDGSL----- 1201  
QY 196 LMLEYQGLIKITLITLTPGQKEGYTELRLPLNQL-----NSIHFSLGENEPLTWSPGYT 251  
Db 1202 -----YGDENYAIRIPPSHWTSIIELR--NKEFKSNPAHKYULA--VDPVSGSLYVS 1253  
QY 252 PIGKNGILGQWISMTAPGGLKETVNSNNQGHFQSANLPLPVYTLMKQVPGAGOR 311  
Db 1254 DYNRRRIY-----RVKSLSGTKDLAGNSEVAG-----TGEQCLP---DEARCGQGK 1299  
QY 312 AIOAEYSYTHNYVGGSGNGIWN-----NKLDNLTGLMTEYVYSGTESRRYK---DK 360  
Db 1300 AIDA--TLMSPRGIAVDKGLMFLVDTMIRKVDQ--NGIISTL--LGSNDLTAVPRLSCDS 1355  
QY 361 EGHDOIVRIE-----RTYNNYHLITSECKQONGYIQTETAYVAILGHNPDSPSQF 412  
Db 1356 SMDVAQVRLPMTDLAVNPMDNSLYVL-----ENNVIILITENHQVSIIA---GRPMHC 1406  
QY 413 QLPKTKETWRSDNSRYSEIETETTPDESGNPLTKVKKDKTKOCLISPTMEYEPAGE 472  
Db 1407 QVPGIDVSLSKLAISHLESASALAIASHTGVLVLTETDEKKINLRQVYTNGBICLLAGA 1466  
QY 473 VDNCPPE-----PYGFTRFVKKLIQTPYDEFPKDP 503  
Db 1467 ASDCDCKNDVNCNCSGDDAAYTDAIILNPSLSLAVADG--TIYADIGNIRIRAVSKNKP 1525  
QY 504 --EKFIYRYSLIGSQSHVTLKIEBHYSATQLNSTL---FOYNTDKSELGRLKQTEC 558  
Db 1526 VLNAPNGYEASPEBQCLYVFNADGIHQYVSLVTEGYLYVFTSTD--NDVTBLI----- 1579  
QY 559 TKGNGKTVSVVHKFTYTKODDPLQOSH-----SITHDNFTIHSQ----- 600  
Db 1580 --DNNGNSLKI-----RRDSSGMPRHLMPDNOIITLTIVGTNGLGLKVSTQWLEGLM 1630  
QY 601 VRSRYTRLFSDDYTKOIVTQMSYDKLGRLLTRTLNSGT-----PYANTLLTYDELANL 654  
Db 1631 TYDNTGLATKSPETGWTTFYDDHDEGRLLNVTRPGVYVSLREMEKSTIIDIEHSNR 1690  
QY 655 QDDNRPPVITTTDVNG-----NQLRNEFD-----GAGRH----- 684  
Db 1691 DDD-----VTVITNLSSVEASVTVVQDVNRSYOLCANNGLLRVWYANGMGISFSEBPHVL 1745  
QY 685 -----VSQC-----LKSDGQKFTYTHIQOYDQGRHHTSYSDYL 721  
Db 1746 AGTTPFTTGRGNISLPMENGINSIEMRLKEQIKGK-VTFGRKLRVNGRNLISI--DYD 1802  
QY 722 TNGR--QOTDPDKVHLSMSKSYDNMGQ-----IANTHWSY-----GVSEK 759  
Db 1803 RNITEKTIYDHRKFTIARIIDYQGRPFMLPSSGLAIVNVSYFENGRLAGLGAGANSE 1862  
QY 760 ITVDPITLTATKQLQNSNNVQTSKEVTTYTPSQOPIQITLPDEAGHLQSCHTLTRDGD 819  
Db 1863 TDID-----KQGRIVSRMFADGK--VMSYSYLDKSWLLAQSORQY----- 1901  
QY 820 RVRKETDALIGCTIYQVDNVRVAVQITLPDGTIVNRKKAAPSTDTLTDIRVNGISLQ 879  
Db 1902 -----LFEYDSSDLRLAVTWP--VAAR--SKSTHTSIGYIR---NITNP 1939  
QY 880 TFDGLSLTQSQDQGRWAWYYSAGNDQCESTVITTPQGFTHQY--QPELDVAVLQVYS 937  
Db 1940 PESNASVIFVDSDDGRIL-----KTSFLGTRGRQVFFYKYGKSLSKSLSEIYYD--S 1985  
QY 938 NEITQOQPSYNDVTGAL--LKAIVAGQSLTPRIYPSGRKLMENINDMKMSYLMWTIRGLN 995



```

Db 1986 TAVT--FGYDETTGVLKVVNLQSGFSCIRYRKIGPL-----VDKQYRFBSEGMVN 2036
Qy 996 GYDLDLTCTIQKISDTHGRVTOIKDSSIKTL-----NYDDL--RHIGS-----QVT 1041
Db 2037 ARPDYT-----YHDSFRIASIKPVISETPLPYDLRYEYBISGKVEHFGFVYIYDIN 2090
Qy 1042 DLATGMLTTTVEPD--GLNREIGRKLCDSGHTLIDIQSGWLKQQLANRIVKLN--GV 1096
Db 2091 QIITTAVMTISKPHDTGRKKEVOYEMFRSLMTMYTQYDSM-----GRVIRELGLP 2144
Qy 1097 LQRTQYVSYSRNLNLYKCDGA-----ECPTDKYGHSLVTQNFYIYGNITACHTT 1149
Db 2145 YANTTKTYD-----YDGGQQLQSVAVNDRPFRX-----SYDLNGLN----- 2182
Qy 1150 PACTEDHATKPAFPDPCQLTEVHTHPRMPPNINLKTDKAGRVNI-----TDNG 1203
Db 2183 -----HLNPNQSVRL-----MP-----LRVDLRDRIRLDDVOYKIDDDG 2218
Qy 1204 -----NTENFTYDTLGRL-----ONGQSVYGYDPLNRLVSOXTDLDCELYRETMLV 1252
Db 2219 YLCORGSDIFBYNSKGLLTRAANKASGWSVOYRDOGRASXTKN--LGHHLQYFYSDLN 2277
Qy 1253 NEVR-----NGEMIRL--RTGETTIAQORASKVLLTGTDSQOS--VILTSKONLS 1300
Db 2278 NPTRITHVYHNSSEITSLVYDLOGHLFAMESSGEYVYASDVTGTPPLAVFSINGLMIK 2337
Qy 1301 QEAYSANGK--HKSTANDASLIGNERADPVSGVTLGNGYSYDPTLMRFHTPD----- 1354
Db 2338 QLOVTAAGEIYYDSNPFOVWIGFHGGLYDPLTCLVHFTQ--RDYDLAAGKMTSPDYTMW 2395
Qy 1355 -----SLSPFGAGGINSYCYLGDPIRNS--DPSGHL-----SMQWMTGI----- 1392
Db 2396 KAVGKEAPF-----NLVYKSNPNLSSELDLKNYVYDVKSWLVMFQJLSNIIPEFPRA 2450
Qy 1393 -----GKGIAGLLLTATGMAIAAG-----IAAIA----- 1421
Db 2451 KMYFVPPPEYELSESQASENGQLIGVOOTTERHQAFMALEGGVITKGLASIEKAGHW 2510
Qy 1422 -SSTTALFAL-----SVTSITISVGALEADAPKASSIIG----- 1459
Db 2511 FATTTPPIIGKIMPAIKEGRVTGVSISAS---ED-SRKVASVLMNAYVLDKMYISIEGK 2566
Qy 1460 -----WVSMGGAAGLA-----ESA-----IKGTLATHLGAFADGE 1493
Db 2567 DTHYFVIGSAGDGLVLTGTTGRKVLDESQVNTVSOPTLLVNCRTFRFNI---EFQY 2622
Qy 1494 NALLKSTSESSRIKMGVT--RSLDREIVRNEGGVYIKDHSRGYTDNFMKGEOAILVHGD 1552
Db 2623 STILLS-----IRYGLTPTLDEKAR-----VLDAQBALGTAMAKEQOKARQD-R 2669
Qy 1553 DGFLYHTEGUK 1563
Db 2670 EGSRLMTEGK 2680

```

# RESULT 10

```

US-10-453-372-136
; Sequence 136, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alcobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-588 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; PRIOR APPLICATION NUMBER: 2003-06-03
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10

```

```

; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuiSeqList version 0.1
; SEQ ID NO 136
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-136

```

```

Query Match 3.1%; Score 272.5; DB 6; Length 2733;
Best Local Similarity 19.5%; Pred. No. 3e-09;
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

```

```

Qy 49 GHIVGNG-NIGPLPLPLSLSPKNTDIFGIFGIFGSLYYDRKNLSLSTGENYK--- 104
Db 1066 GHIFQSFQSPASPIASTFW--DKTDA-----YGQRYGASDAVASV--GEYETCP 1113
Qy 105 --VLETDKTVKLOOKLD--NL--RPEK-----DLKENCYRIHK--SGDIEVLTG-- 147
Db 1114 SLIMKERTLLDGLFELDPENLGSMSLDKHLILVSG--ILHKIGENQPLTQPAII 1170
Qy 148 -----FNNAPDLKVP-----KKLNPAGHAIYIDWNEATQPLNRIYDLDGDIP 195
Db 1171 TSIMGNRRRSISCPSCGNGLABGNKLLAPVALAVGIDGSL----- 1210
Qy 196 LNLLEYQGLIKTLTLPQKGEGRTELRLNQL-----NSINFLSGENPLTWSGYT 251
Db 1211 ----YGDPNYIRIRFPSSNVTSTLELR--NKEFKSNPAPKHYIA--VDPVSGSLYVS 1262
Qy 252 PIGNGILQWITSMTAPGLKETVYVSNNOGHHPQSANLPLPYVTLMKQVPGAGOP 311
Db 1263 DMSRRRY-----RVKSLSTGDLAGNSVYAG-----TGBQCLP--DEARCGGCK 1308
Qy 312 AIOAESYTSHTNYVGGSGNIN-----NKLDNLYGMLTEYNYSSTESRRYK---DK 360
Db 1309 AIDA--TLMSPRGIAVDKNGIAMFVDATMIRKVDQ--NGIISTV--LGSNDLTAVRPLSCDS 1364
Qy 361 EGHDIYRIR-----RTYNNYHLLTSECKQONGIOTTEFAVYALIGHNDSQPSOF 412
Db 1365 SMDVAQVRLFWPTDLAVNPMDNSLYVL---BNNVILRLTTEHNOVSLA---GRPMHC 1415
Qy 413 QLEPKTETWRSADNSYRSEITETTPDESQNPITKVKIDKKTOKIISPSTHMEYRPAGE 472
Db 1416 QVPGIDYSLSKLAHSLBSASAIAISHTGVLYITETDEKIRMLQVTTNGEICLLAGA 1475
Qy 473 VDNCPPE-----PYGTFRVKKIIQTPYSEFKDP 503
Db 1476 ASDCDKNDVNCNCYGDDAYATDALNSPSSLAVAPDG--TIYIADIGNIRIRAVSNGKP 1534
Qy 504 --EKFYQYRSLIGSGSHVTLKIEBHYSATQOLNSTL---FOYNDKSLGGLLKQTEC 556
Db 1535 VLNAFQYENASFGEBELVFNADGIHQYTVSLVGBEYLNFYSTD--NDVTBLI 1588
Qy 559 TKGENGTYSVHKFTYTKODDTLQOSH-----SITTHDFTIIRSQ----- 600
Db 1589 --DNNGNSLKI-----RDSGGMRRHLLMPNOIITLTVGNGGLKXVSTQNLBGLM 1639
Qy 601 VRSRYTGRFSDTDKQIVTQMSYDKLGRLLTRTNSGT-----PYANTLYDYELNLT 654
Db 1640 TYDNGTGLATKSDDEGWTFFYDYDHEGRGLTNVTRPVGVTSLHREWEKSIITIDIENSNR 1699

```

```

Qy 655 QDNRPPVITTTDVG-----NOLRNEPD-----GAGR-----684
Db 1700 DDD-----VTVITLSSVEASVTVQDQVRNSYQLCNNGLTARVYANMGISFSEPHVL 1754
Qy 685 -----VSQC-----LKSDGDKFTYHTQOYDEGRHSTYSDDL 721
Db 1755 AGTTPTTRGCNISLPMENGINSIEMRLRKEQIKGK-VTIFGRKLRVGRMLSI--DYD 1811
Qy 722 TNGR-QOQDPPKVLMSKSYDNMGQ-----IANTHSY-----GVSEK 759
Db 1812 RNITEKTYDHRKFTLTIYDQGRPPLMLPSSGLAIVANSYFFNGHLAIGLQANSER 1871
Qy 760 ITVDPTLTATKOLQSNNSNVQGEVTTYTPSOQPIQITLPDEAGHLQCHTLTRDGD 819
Db 1872 TDID-----KQGRIVSRMFADGK-VMSYGLKSMVLLQSQROQ-----1910
Qy 820 RVRKETDAIGCCTIYQVNNRNVIOITLPDGTYNRKKAPESTDTLTDIVNGISLQOQ 879
Db 1911 -----LFEYDSDRLLAVTWPS--VARH--SMSHTSICGIR--NIYNP 1948
Qy 880 TFDGLSRLTQSDQDGRVWAVTYYSAGNDQCPSFTVITPDGQFIHQY--QPELDAVLQVAS 937
Db 1949 PESNAAVIFDYSDDERIL-----KTSFEGTRQVFPYKYGKLSKLSIYVD--S 1994
Qy 938 NEITQPSYNPVTGAL--LKAVAEQSLTPYPSGRLKMEINDMKMSYLMTLRGLN 995
Db 1995 TAVT--FGYDETTGVLKVVNLQSGGFSCTIRYRKIGPL-----VDKQIRFSEBGMVN 2045
Qy 996 GYTDLTGITQKISRPTHGRVTOIKDSSIKITL-----NYDDLN--RHIGS-----QVT 1041
Db 2046 ARFDT-----YHNSFRIASIKFVISETLPVLDLYRYDEISGKVEFGKGVLYYDIN 2099
Qy 1042 DLATGHMLTTTVEPD--GLANREIGKLCDSGHTLIDIQSWLKTQOLANRIVKLN--GV 1096
Db 2100 QITTAWTLKSKHEDTHGRIRKEVQYEMFRSLMTMTVUYDSM-----GRVYKELKGP 2153
Qy 1097 LQRTQOYSYDSRNLNOYKCDGA-----ECPTDKYSHSVTONFYVDIYGNITACHTT 1149
Db 2154 YANTTKXYD-----YDGDQQLQSVAVNDRPWTRY-----SYDLGNL-----2191
Qy 1150 FADGTEDHATKFNPTPCQLTENVHHTHPMPDNIRLKYDKAGRVINI-----TDNHG 1203
Db 2192 -----HLNLPNGSVRL-----MP-----LRYLDRKITLQDVOYKXIDDDG 2227
Qy 1204 -----NTENFTYDVLGRL-----ONGQSVYGYDPLNRLYSOKTDLDCELYYRETLV 1252
Db 2228 YLCQSGSDIFBYNSKGLLTRAYNKASGWSVOYRIDGVRBRASYSKTN-LGHLLQTFYSDLH 2286
Qy 1253 NEVR-----NGEMIRLL--RTGETTIAOQASKVLTLTDSQOS--VILTSDKONLS 1300
Db 2287 NPTRLTHVYHNSNSBITSLAYYDLQGLHLPAMESSGSEYVVASDNTGTPLAVFSINGLMIK 2346
Qy 1301 QEASVAYGK--HKSTANDASILGNGERADPVSGYTHLGNGYRSVDPILMFPHPD-----1354
Db 2347 QLOVAYGEIYYDSNPDQOMVIGFPGGLYDPLTKLVHFTQ--RDVDVLAGMTSPSDYTMW 2404
Qy 1355 -----SLSPFAGAGINPYSCIDPINS--DESGHL-----SMQAWTGI-----1392
Db 2405 KNVGEPRAPF-----NLTMFKSNPDLSEBDAKNTVTVDKSVLVMFGQLSNITIPFPRA 2459
Qy 1393 -----GMGIAGLLTITATGMAIAAGG-----IAAIA-----1421
Db 2460 KMVFVPPEYELSESQASENGQLITGVQOTTERHNOAPALBEGQVITKKLHLSIREKAGHW 2519
Qy 1422 -STSTALAFGL-----SVTSDITSVSGALEASFKASIIIG-----1459
Db 2520 FATTPIIGKIGMFAIKGGRVTVGSIIAS--ED-SRKVASVLMNAVAYLDMKMYSIDGK 2575
Qy 1460 -----VWSMGGAAGLA-----ESA-----IKGQTKLATHAGAFAPDGE 1493
Db 2576 DTHYVVKIGSADGLVLTGTTIGKAVLSSGVAVVTVSOTPLLVNGTRTFNTI--EYQ 2631
Qy 1494 NALLKSTSESSRIKWGVT-RSLDREIVANEQOVIKDHSGRYTDFNMKGQAOAILVHGD 1552

```

```

Db 2632 STILLS-----IRYGLTPDPLDBEKAR-----VLDAQORALGTAMAKEQKARDG-R 2678
Qy 1553 DGFVHTGK 1563
Db 2679 EGSRLWTEGK 2689

RESULT 11
US-10-453-372-142
; Sequence 142, Application US/10453372
; Publication No. US2006003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsebrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
; FILE REFERENCE: 21402-569 A
; CURRENT APPLICATION NUMBER: US/10/453.372
; PRIOR APPLICATION NUMBER: 2003-06-03
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Cnaseqdist version 0.1
; SEQ ID NO 142
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-142

Query Match 3.1%; Score 272.5; DB 6; Length 2733;
Best local similarity 19.5%; Pred. No. 3e-09;
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

Qy 49 GHIVGNG-NIGPTPLPLTSLSPLNKTDIGFGIGFPGSVYDRKNSLSLSTGENYK---104
Db 1066 GHLLQKSFQASPNLASFITM--DKTDA-----YQGRVGLSDAVVSV--GEYETGCP 1113
Qy 105 --VLETDKTVLQOKLDD--NL--RPEK-----DLKENCYRIIRK--SGDIEVLTG-----147
Db 1114 SLIMERTKTLLOGFELDPNSLIGGWSLDKHILNVKSG--ILHKVGENGFLTQOPALII 1170
Qy 148 ----FNNAFDLKYF-----KKLLNPAGHAIIYDMNEATQPLRLNRIYDDIDGDHDP 195
Db 1171 TSIWNGRRRRSISGCPGCLAEKGLAPVALAVAGIDGSL-----1210
Qy 196 LAMLEVQGLIKTILTLTPGQKEGRTLRFLNRLD-----NSIHNPISGENPLTMSFGYT 251
Db 1211 ----YGDENVYIRIRIPSRNVTSILBLR--NKEFKSNNDPAHKYILA--VDPVSGSLYVS 1262
Qy 252 PIGKNGILGOWITSMTPAGLKETVANSNNQGHFPQASANTPLPYVTLAKQVPGAGOP 311
Db 1263 DTNSRRITV-----RVKSLSGTKDLAGNSVYVAG-----TGEQGLTF--DEARGDGGK 1308
Qy 312 AIOAEYSYTSNHYVGGSGNGIWN-----NKLNLVGLMTVEYVYGGTSESRRYK---DK 360

```

```

Db 1309 AIDA--TLMSPRGIAVDKNGLMFVDAWTMRKVQD--NGIISTL--LGSNDLAVAPLSGDS 1364
Qy 361 EBGHDIYRIR-----RTYNNHLLTSECKQNGIOTTERAYVALIGHNPSQSOF 412
Db 1365 SMDVAQRLTEWPTDLAVNPMDSLYVL-----ENNVLIRITEMQVSIIA--GRPMHC 1415
Qy 413 QLPKTKETRSADNSYRSEITETTFDESQNPFLKVKDKKTKOKIISPTMHEXPAGE 472
Db 1416 QVGEIDYSLKLAHSLBESASALASHITGVLYTTERDEKKINRLROVTTNGEICLACA 1475
Qy 473 VDNCPPE-----PYGFTFRVYKIIQTPYDSFKDDP 503
Db 1476 ASDCCKNDVNCYSGSDAAYATDALINSPSLAVAPDQ--TIYADIGNIRIRAVSNKP 1534
Qy 504 --EKFTIYRSLIGSOSHVLTKIEBRYSATQLANSTL--FOYNTKSELGRLKQTEC 558
Db 1535 VLNAPNVEASPEOEIYFNADGIHQYTVSLVGEYLYNFYSTD--NDVTELI----- 1588
Qy 559 TKGNGKTVSVHKTFTYTKODDTLQOSH-----SITHDNFTIHSQ----- 600
Db 1589 --DNNGSLKI-----RRDSGMPRHLMPDNQIITLVGTVGGLKVVSTOMLEIGLM 1639
Qy 601 VRSRYTSLPSDPTDKOIVTQMSYDKLGRLLTRTLNSGT-----PYANTLYDYELANTL 654
Db 1640 TYDQNTGLATKSDQETGWTTFYDYDHGRLTNVTRPGVTVSLHREMEKSTIIDISNR 1699
Qy 655 QDDNRFPVYITTTDVNG-----NOLNREF-----GAGRH----- 684
Db 1700 DDD-----VTVITNLSSVEASVTVVQOVNRSYOLCANNGLRVWYANGMIGISFSEPHVL 1754
Qy 685 -----VSQC-----LKSDGDKFYTHITQOYDQGHSTYSDDL 721
Db 1755 AGTITPTIGKCNLSLPMENGLNSTEMRLRKEQIKGK--VTIFGRKLVRHGNLSI--DYD 1811
Qy 722 TNGR--QCTDPDKVHLNMSKSYDMNGQ-----IANTHNSY-----GVSEK 759
Db 1812 RNIRTEKIDYDHRKFTLRITIDYQGRPLMLPSSGLAANVSYPFNRLAGLQKANSER 1871
Qy 760 ITYDPTLTLATKQLOQSNVNVQSKVYTTTPSQPIQITLFDAGHLQSCHTLTRDQMD 819
Db 1872 TDD-----KQORIVSRMFADK--VMSYSYDLKSNVLLQSGORQ----- 1910
Qy 820 RVRKETDAIGCCTYYOYDNVNRVQITLPPDGTIVNRKXAPSPDTLTDIRVNGISLQOQ 879
Db 1911 -----TFEYDSSDLAATMPS--VAKH--SMKSTHSGYR--NITYNP 1948
Qy 880 TFDGLSLTQSDGGRVWAAVYYSAGNDQCSVTITPDQFPIHQY--QPELDAVLQVAS 937
Db 1949 PBNASAVIPYSDDGRL-----KTSPLGGRQVFFYVYGLSKLSSELYVD--S 1994
Qy 938 NEITQOSYNPVTGAL--LKAVAEGOSLTPYYSGRLKKNENINDMKMGSLYTLRGLN 995
Db 1995 TAVV--FGYDETGVLMQVNLQSGFSCITRYRKIGPL-----VKKOYRFESEGMVN 2045
Qy 996 GYTLTGTIQLKISHDTHGRVLTQIDSSIKTLL-----NYDDLN--RHISG-----QVT 1041
Db 2046 ARPDYT-----YHNSFRLASIPVYSETPRLPYDLRYRDEISKVEHFGVGYIYDIN 2099
Qy 1042 DLATGMLTTLVTEPD--GLNREIGRKLDSGHTLIDQOSVLTQOLANRIVKLN--GV 1096
Db 2100 QIITTAAMTSLSKHFDTHGRIKREVQEMFRSLMMYMTQYUYSM-----GRYIKRELKGR 2153
Qy 1097 LQRTQYSSYSRRLNLYKCDGA-----ECPLDKYGHSLVTQNFYUUYDIYNTACHTT 1149
Db 2154 YANTTKTYD-----YDGGGQLOQSAVANDRPMTRY-----SYDLNGLN----- 2191
Qy 1150 FADQTEHATFKRANPTDPCQTEVHTHPRMNPINILKYDKAGRVINI-----TDNHG 1203
Db 2192 -----HLNPNRSVRL-----MP-----LRIDLRLRITRLDQVYKIDDDG 2227
Qy 1204 -----NTENFTYDTLGLRL-----QNGQGSVYUQDPLNRLVSQKTDPLDCELYURETTLV 1252
Db 2228 YLCQSGSDIFRYNSKGLLTRAUNKASGMSVQYVYDQGRBASUKTN--LGNHLQYFYSDLN 2286

```

```

Qy 1253 NEVR-----NGEMIRLL--RTGETIINOGRASKVLNLTGDSQOS--VILTSKONLS 1300
Db 2287 NPIRITHVNHNSSETLSLYDLOGLHFNAMESSGEERYASDNTGTPPLAVFSTINGIMIK 2346
Qy 1301 QBAVSAVGK--HKSTANDASILGNGERADPVSGVTHLGNNGYSYDPTLRFPD----- 1354
Db 2347 QLOVTAIVGELIYDSNPDQFQVIGFHLGIDPLKLVHFTQ--RDYDLAGRMTSPDYTMW 2404
Qy 1355 -----SLSPFGAGGINDPYSICLADPPINR--DPBGHL-----SQOAWGTI----- 1392
Db 2405 KNVGEKBPAP-----MLYMKSNPNPLSSELDLKNYVYDVVSMVMEFQLSNIIIGPPRA 2459
Qy 1393 -----GMGIAGLLTLATGMAIAAG-----IAATA----- 1421
Db 2460 KMTFVPPPYLSSQASENQULTTVOQOTTERNOAFMALEGOVITTKGLASIREKAGHW 2519
Qy 1422 -STSTLAFGAL-----SVTSDITSYSGALEDAFPKASILG----- 1459
Db 2520 FATTPTIGKINFALKEGVTGVSIS--ED--SRKVASVLNNAVYLDKXHYISBK 2575
Qy 1460 -----WVSMGMAAGLA-----ESA-----IKGTGLATHLQAFADGB 1493
Db 2576 DTHYFVYIGSADDLVTIAGTIGRKVLESGVNTVSGPTLLVNGRTFRFTNI--EFOY 2631
Qy 1494 NALLKTSSESRKMGVT--RSLDRBYRNEEGVYIKDHSRGYDNTFKGKEQALVYHGDK 1552
Db 2632 STLLS-----IRYGLTPTLDEKAR-----VLDQARQALGTAMAKEQKARDG-R 2678
Qy 1553 DGFYHTEGK 1563
Db 2679 EGSRLWTEGK 2689

```

```

RESULT 12
US-10-453-372-146
; Sequence 146, Application US/10453372
; Publication No. US2006003323A1
; GENERAL INFORMATION:
; APPLICANT: Alcobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 146
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-146

```

Query Match 3.1%; Score 272.5; DB 6; Length 2733;  
 Best Local Similarity 19.5%; Pred. No. 3e-09;  
 Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;

QY 49 GHVYNG-NIGPTPLPLTSLSPKLTIDIGRIGNFGLSYDRKNSLSTGENYK--- 104  
 DB 1066 GHLEFQKSFQASPNLSTFTW---DKTDA-----YQORYVGLSDAVAVSV--GREYETCP 1113  
 QY 105 --VLETKTVAGLQOKKD--NL-----RPEK-----DLKNCYRIHK--SGDIEVLG----- 147  
 DB 1114 SLIMEKRTALLQGFELDPNSLWGSMDKHHILNVKSG--ILHKGTGENOPLTQQAII 1170  
 QY 148 ----FNNAFPLKVP-----KCLINPAGHAIIYIMNFEATOPRLNRIYDDLGDHDP 195  
 DB 1171 TSIINGNRARRISCPSCNGLAEGNKLLAPVALAVIGDSL----- 1210  
 QY 196 LNLNFGGLKTLITLTPGQKEGYRTERFLNROL---NSIHNFSICNENPLTWSFGYT 251  
 DB 1211 ----YVGDFFNYIRIIPSRNVTSLLEIR--NKSEFKISNNPAHKYULA--VDPVSGSLYVS 1262  
 QY 252 PIGKNGILGQWITSMTPAPGLKETVNSNNNGHFPQSANLPLVPLVYTLMKQVPGAGOP 311  
 DB 1263 DTNSRRYI----RVKSLSTGKIDLANSSEVAG-----TGEQCLPF---DEANCGSGK 1308  
 QY 312 AIOAEYSYTHNVYGGSGNGIMN-----NKDLNLYGLMTEYNGSTESRRYK-----DK 360  
 DB 1309 AIDA--TLMSPRGIAVDKNGLMFYDAMIRKVDQ--NGIISTL--LGSNDLPAVAPLSCDS 1364  
 QY 361 EGHQOIVRIE-----RTYNNHLLTSECKQONGYIQTETRYVALIGNPSPQSOP 412  
 DB 1365 SMDVAQVRLWPRTDLAVNPMNLSLYL-----ENNVLIRITENHOVSIIA---GRPMHC 1415  
 QY 413 QLPKTKETWMSADNSYRSBITETTFDESGNPLFKVIKDKTKOKIISPTEMEYPAGE 472  
 DB 1416 QVPEIDISLGLAIHSALESASALAIHSHTGVLYITTEDEKIKINLRQVTNGEICLAGA 1475  
 QY 473 VDNCPPE-----PYGFRFYVKIIQTPYDESEFROP 503  
 DB 1476 ASDDCNDVANCYSGDDAATDALIINSPSLAVAPDG--TIYADLGNIRIRAVSNKP 1534  
 QY 504 --EKFIQRYRLISSQSHVTLKIEBRHYSATQLANSTL---FOYNTDKSELGRLLKQTEC 558  
 DB 1535 VLANFNQVEASPEQELVYENADGIIHQYTVSLVGEVLYNFTYSTD--NDVTELI----- 1588  
 QY 559 TKGENKTYSVVHKFTYKODDTLOQSH-----SITTDNFTHRSQ----- 600  
 DB 1589 --DNNGNSLKI-----RDSGSGPRLLMPDNOIITLVGTNGGLKVSTQMLEGLM 1639  
 QY 601 VRSRYTGLFSDDTDKDITVQMSYDKLGRLLTRTLNSGT-----PYANTLTYDELANL 654  
 DB 1640 TYDGNLTGLATKSDBTGWTTFYDDHEGRLTNVTRPTGVVTSLHREMEKSIITIDIENSNR 1699  
 QY 655 QDDNRPPVITTTDVNG-----NOLRNEFD-----GAGR----- 684  
 DB 1700 DDD-----VTVITLMSVEASYVVOVDVNSYOLCNGGLRVWYANGMGSFSEPHVL 1754  
 QY 685 -----VSQC-----LKDSGDGKFYTIHQYDEQGRHHTSTYSDDL 721  
 DB 1755 AGITPTTGRGNISLPMENGLNSTEMRLKXQIKGK-VTIFGRKLRVGRMLSLI--DYD 1811  
 QY 722 TNGR-QQTDPPKVLHLSMSKSYDNMGQ-----IANTHMSY-----GVSEK 759  
 DB 1812 RNIRTEKYIDHRRKFTLRITIDQVGRPLMLPSSGLAIVANSYFENGRLAGLQGANSEK 1871  
 QY 760 ITUPBITATKQLOLSNNVOTGEVTTYTPSQOPITQITLEDAGHLOSHCHITLRDWD 819  
 DB 1872 TDID-----KQGRIVSRMPADGK--VWSYSTLDKSMVLLLOQSQOY----- 1910  
 QY 820 RVKRETDALIGCTIYQDNYNRVIOITLPDGTIVRKAPFSTDTLIDIRNGISLQO 879  
 DB 1911 -----IPEYDSBRLAVTMPG---VARH--SMSTHSTIGYR-----NINYP 1948  
 QY 880 TFDGLSRLTGOSDGGRVWAVYYSAGNDQCPSTVITPPDQFIHYOY--QPELDVAVLQVAS 937

DB 1949 PESNASVIFPYSDDGRIH-----KTSFLGTRGVFPYKXGLSLSSELYVD--S 1994  
 QY 938 NEITQGFVNPVYUGAL--LKAVAEQSLPFIYPSGRLLKMEINNDKMSGYTLTRGLEN 995  
 DB 1995 TAVY--FGYDETTGLMVMVLOSGGFSCTIRYKIGPL-----YDKQIYRSEBGMVN 2045  
 QY 996 GYVTLGTITOKISRDTHGRVTOIKDSIKTTL-----NYDDLN--RHIGS-----QVT 1041  
 DB 2046 ARPDYT-----YHDSFRASIKPVISETPLPVLDYRYDEISGKVEHFGKFGVIYYDIN 2099  
 QY 1042 DLATGHMLTTTVEFD--GLNREIGRLCDSSGHTLDDIQSWLTKTQOLANRYKLN--GV 1096  
 DB 2100 QIITVWMTLSKAFDTHGRVKEVOYEFRLMVMYVQYDSM-----GVYIKRELKGP 2153  
 QY 1097 LQRTQSYDSRRLNQKDDGA-----ECPTDKGHSIVONFTYDIGNITACHTT 1149  
 DB 2154 YANTTKTYD-----YDGGQLOQSVAVNDRPTWRT-----SYDLANGN----- 2191  
 QY 1150 FADGTEDHATFKFANPTPCQLTEVHTHPDMPDNRLKYDKAGRVINI-----TDNHG 1203  
 DB 2192 -----HLNPNQSVRL-----MP-----LRYDLDRITRLGDVQYKIDDDG 2227  
 QY 1204 -----NTENFTYTLGRL-----QNGQGSVGYDPLNRLVSOQKTDPLDCELYIREMTLV 1252  
 DB 2228 YLCORGSDIPEYNSKGLTTRAYNKASGMSVOYRYDVGRAASYKTN--LGHRLQFYSDLH 2286  
 QY 1253 NEVR-----NGEMIRLL--RTGETIIAQBASKVLITGTSQOS--VILTSDKONLS 1300  
 DB 2287 NPRTIRITHVNSSETISLYYDLOGLHFMESSGEBEYVYASDNTGTPPLAVPSINGLMIX 2346  
 QY 1301 QEAYSAYGK--HKSTANDASILGNGERADPVSGVTHLNGYRSYDPLMRFHTPD----- 1354  
 DB 2347 QLOQYTAGELIYDQSNPDFOVWIGFHGGLYDPLTKLVHFTQ--RDYDVLAGRMTSPDYTMW 2404  
 QY 1355 -----SLSPGAGINPYSICLQDPINRS--DPSGHL-----SQWAMTGI----- 1392  
 DB 2405 KNYGKEPAPF-----MLYMKFSNNPLSSELDLKNYVYDVXSWVMEFPQLSNIIIPGPRA 2459  
 QY 1393 -----GMGIAGLLTLATGGMALIAAG-----IAAIA----- 1421  
 DB 2460 KMTFVPPPEYLSQSASENQOLITVQOTTERNOAMALEGVIITKGLASIEKAGHW 2519  
 QY 1422 -STSTALAFAL-----SVTSDITSIVSGALEDASPKASIIIG----- 1459  
 DB 2520 FATTPPIIGKIMPAIKEGVTTGVSSIAS--ED--SRKVASVLNNAVYLDKMHYSIEGK 2575  
 QY 1460 -----WVSMGGAAGLA-----ESA-----IKGTGLATILGAFADGE 1493  
 DB 2576 DTHYFVYIGSADGDLVTLGTTIGRKVLESQVNVTVSQPTLLVMGRTIRFTNI--EFQY 2631  
 QY 1494 NALLKSTSESRIRKMGVT--RSLDEIYRNEGVYIKDSRGYTDNFMKGEOALLVHGDK 1552  
 DB 2632 STLLLS-----IRYGLTPTLDEKAR-----VLDQAORALGTMAAEQOQARBOG-R 2678  
 QY 1553 DGFLYHTEGNK 1563  
 DB 2679 EGSRLMTEGER 2689

RESULT 13  
 US-10-453-372-150  
 ; Sequence 150, Application US/10453372  
 ; Publication No. US2006000323A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alcobrook, et al.  
 ; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
 ; FILE REFERENCE: 21402-589 A  
 ; CURRENT APPLICATION NUMBER: US/10/453, 372  
 ; CURRENT FILING DATE: 2003-06-03  
 ; PRIOR APPLICATION NUMBER: 09/789390  
 ; PRIOR FILING DATE: 2001-02-23  
 ; PRIOR APPLICATION NUMBER: 60/185967



QY 1460 -----WVSMGMAAGLA-----ESA-----IKGKTATATGAFADGE 1493  
DB 2576 DTHYFVKGISADGDLVLTGTTIGKVLBSGVNTVVSQPTLLVNGKTRFRFINI-----EFQY 2631  
QY 1494 NALIKSTSESRIRKMGVT--BSLDREIVRNEBGVYIKDHSRGYTDNFMKGEGQALVHGDK 1552  
DB 2632 STILLS-----IRYGLVPTDLDEKAR-----VLQARQBALGTMAKEQQAADG-R 2678  
QY 1553 DGFLYHTEGAK 1563  
DB 2679 EGSRLWTEGK 2689  
RESULT 14  
US-10-453-372-154  
; Sequence 154, Application US/10453372  
; Publication No. US2006003323A1  
; GENERAL INFORMATION:  
; APPLICANT: Alabrook, et al.  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
; FILE REFERENCE: 21402-589 A  
; CURRENT APPLICATION NUMBER: US/10/453,372  
; PRIOR APPLICATION NUMBER: 2003-06-03  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/185967  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 09/823187  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195792  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/839446  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/199476  
; PRIOR FILING DATE: 2000-03-25  
; PRIOR APPLICATION NUMBER: 09/863776  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: 60/208263  
; PRIOR FILING DATE: 2000-05-31  
; PRIOR APPLICATION NUMBER: 09/939398  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: 60/227800  
; Remaining Prior Application data removed - See file wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1609  
; SOFTWARE: Curation version 0.1  
; SEQ ID NO 154  
; LENGTH: 2733  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-453-372-154  
Query Match 3.1%; Score 272.5; DB 6; Length 2733;  
Best Local Similarity 19.5%; Pred. No. 3e-09;  
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;  
QY 49 GHYANG-NLDPPTLPLTSLYSPLNTDIFGFGFGLGSYVDRKKSLSSTGENYK--- 104  
DB 1066 GHLQKSFQASPNLSTFIW---DKTDA-----YQGRVYGLSAVVSV--GFEYETCP 1113  
QY 105 --VIETDKTVKLOQKLD--NL--RPEK---DLKENCYRIHK--SGDIEVLTS----- 147  
DB 1114 SLIMEKRTALLQGEFLDPSNLGWSLKHHLNANKSG---ILHKGTEBNGFLTQOPAI 1170  
QY 148 ---FNNAFDLKV-----KKLNPAGHAIYIMNFATOPRLNRIYDDLDGDHIF 195  
DB 1171 TSMINGRRRISPCSCNGLAEGNKLLAVALAVGIDGSL----- 1210  
QY 196 LNLNLEQGLIKITLLPFGQKGYRTFLNROL-----NSIHNSLGNENPLTWSFGYT 251  
DB 1211 ---YVDENYIRIFPSRNVTSILELR--NKEPKHSNNPAHKYLLA--VDPVSSSLYVS 1262  
QY 252 PIGKGIIGQWITTSWTAPGGLKETVNVGNNNGHHFPOSANLPLVPLRYUTLAKQVVGAGAP 311

DB 1263 DTSRRRIY-----RVSLSGTKOLAGSEVAVG-----TGEQCLP---DEARCGGK 1308  
QY 312 AIOAEISYSHNVYGGGNGINW-----NKDNLGLMTENVYSGTESRRYK---DK 360  
DB 1309 AIDA--TLMSPRGIAVDKDKLMTFVDATMIRKQD--NGIISTU--LGSNDTLAVAPLSCDS 1364  
QY 361 EGHDOIVRIE-----RTYNNVHLITSECKOONGYIOTETRAYVAIIGHNFPDSQSF 412  
DB 1365 SMDVQVRLBWPDPDLAVNPMDNSLYVL-----ENNVILRIITENHOVSIIA---GRPMHC 1415  
QY 413 QLPKTKETWRSADNSYRSEITETPEDESGNPPLTKYIKDKKTKOKIISPTSHMEYYPAGE 472  
DB 1416 QVPGIDVSLKLAHSLAESASAIASHTGVLTYTETDEKINRLRQVTTNGEICLLAGA 1475  
QY 473 VDNCPPE-----PYGTRFVKKIIOFPYDESEFDDP 503  
DB 1476 ASDODCKNDVNCNCGSDDAYATDAILNSPSLAVAPDG--TIYIADGNIIRIAPVSKXP 1534  
QY 504 --EKFIQYRSLIGSGHVTLKIEERHYSATQLLNSTL---FOYNTDSEGLRLKQTEC 558  
DB 1535 VLNAPNOYEAASRQDELVFPNADGIHQYVSLVTGSLNFTYSTD--NDVTELI----- 1588  
QY 559 TKGNGKTVSVHKKFTYTKODDTLQOSH-----SITHDNFTIHRSQ----- 600  
DB 1589 --DNNGNSLKI-----RDSGMPRHLLMPDNOIITLTVGTNGGLKVVSTQMLBGLM 1639  
QY 601 VRSRYTGLRPSDDTDKDIYVQMSYDKLGRLLTRTANGT-----PAANTLYVEILNLT 654  
DB 1640 TYDQNTBLATKSGSETGWTFFYDHDGKRLTNVTRPGVVTSLHREMEKSTIIDIEANSR 1699  
QY 655 QDDNRPPFVITTPDVNG-----NQLRNEFD-----GAGRH----- 684  
DB 1700 DDD-----VVIYNLSVVASVYVQDVANSYGLCANNGLTRVWYANGMGSFSEPHVL 1754  
QY 685 -----VSQC-----LKSDGDKFTYTHTQOYDQGRHHTSYSDYL 721  
DB 1755 AGTTPFGICNLSLPMENGLNSIEMRLRKEQIKGK--VTIFGRKLRYHGRMLSI--DYD 1811  
QY 722 TNGR--QOTDPPKHLMSKSYDNMGQ-----IANTMSY-----GVSEK 759  
DB 1812 RNIRTEKIYDHRKFTLRIIYDOGRFPLMPSSGLAIVANVSYPFNGRLAGLQANSEK 1871  
QY 760 ITVDPITLTATKQLOQSNNNVQSGEVTYTPSQOPIQITLFDAGHLQSGHTLTRDGD 819  
DB 1872 TDID-----KQRIYSRKFAQDK--VMSYSYLDKSVLLQSORX----- 1910  
QY 820 KVRKETDAIGQCTIYQYDYNRNVIQITLPDGTIVNRKYAPSTDTLITDIRNGISLQOQ 879  
DB 1911 -----IFBYDSSDRLLAVTMPs---VARH--SMSTHTSIGYIR-----NIYNP 1948  
QY 880 TFDGLSLTQSGDGRWATYTSAGNDQCESTVITPPGQFIHQY--QPELDVAVLQVAS 937  
DB 1949 PESNASVIFDYSDDGRIL-----KTSFLGTGRQVYFYKGLSKLSSEIYD--S 1994  
QY 938 NEITQGSYNPVGTAL--LKAVALGOSLTPYIPSGRLKMEINDMKMSYLAWLKLEN 995  
DB 1995 TAVT--FGYDETTEVLKMNVLQSGGFCGTIRYKIGPL-----YKQIIRFSEBGVN 2045  
QY 996 GYTDLTGTQIKISDTHGRVTOIKDSSIKTTL-----NYDDLN---RHIGS-----OYT 1041  
DB 2046 ARPDYT-----YDNSFRIASIKRVISETPLPDLYKRYDISGVEHFGFGLVYDIN 2099  
QY 1042 DLATGHMLTTEVD--GLAREIKRKLCDSGHLLDIOQSLKTKQOLANRLVKLN---GV 1096  
DB 2100 QITTAWTLISKHDPGRIGRIEVOYEMFRSLMYMTVOYDSM-----GRVYKELKGP 2153  
QY 1097 LQTEQYSYDSRNLNLYQKCDGA-----ECPDYKGSIVTQNFYDIYGNITACHTT 1149  
DB 2154 YANTTKTYD-----YDGGQQLQSVAVNDRPTWRY-----SYDLANGNL----- 2191  
QY 1150 PADGTEDHATFKFANPDPQCLTEVNHHTHPRMPNINILKYDKAGRVINI-----TDNHG 1203

```

Db 2192 -----HLINPNSVRL-----MP-----LRYLDRIRTRLGVQKIDDDG 2227
Qy 1204 -----NTNFTYDTLGLT-----QNGQSVYGVDPPLRLVSOQTDPLDDELYRRTMLV 1252
Db 2228 YLCORGSDIPEYNSKGLITRAYNKASGWSVOYRDGGRASXYXTN-LGHHLQYFYSDLH 2286
Qy 1253 NEVR-----NGEMIRLL--RTGETIIAQPARSKVLTGTSQOS--YILTSKQNTLS 1300
Db 2287 NPIRITHVNHNSSEIRSLVYDQGHLFAMESSSGEYVYASDNTGTPPLVPSINGLMIX 2346
Qy 1301 QEAYSAYGK--HKSTANDASILGNGERADPVSGVTHLGNGYSYDPTLMRFHTPD---- 1354
Db 2347 QLODYTAAGEIYYDSNPFQWVIGFHGGLYDPLTGLVHFTQ--RDYDLVAGRWTSPTYTM 2404
Qy 1355 -----SLSPGAGINPYSCLDPIRNS--DPSGHL-----SQWAMTGI----- 1392
Db 2405 KNVGKEBAPF-----NLMPKSNPNLSSELDLKNYVDVKSWMVWFQOLSNIIPGPRA 2459
Qy 1393 -----QMGJAGLLLTATGMAIAAG-----IAATA----- 1421
Db 2460 KMYFVPPPYLSSESQASENQOLITGVQOTTERHQAEMALEGVITTKLAASIREKAGW 2519
Qy 1422 -STSTTALAFCAL-----SVTSDITSIVSGALEDASPKASITG----- 1459
Db 2520 FATTPTPIGKIMFALKEGRVTTGVSSIAS--ED-SRKVASVLNNAVYLDKMHYSIEGK 2575
Qy 1460 -----WVSMGGAAGLA-----ESA-----IKGCTLATHLGAFADGE 1493
Db 2576 DTHYFVAGISADGLVTLTGTTIGRKVLSEGVNVTVSQPTLLVNGRTFRFTNI--BFQY 2631
Qy 1494 NALLKSTSESSRIKMGVT--RSLDEIYRNEEGVYIKDHSRGYTDNFMGKEQALILVYGDK 1552
Db 2632 STILLS-----IRYGLTPTDLDEKAR-----VLDQARALGTAMAKEQOKARDG-R 2678
Qy 1553 DGFLYHTEGK 1563
Db 2679 EGSRLMTEGK 2689

```

RESULT 15  
US-10-453-372-116  
Sequence 116, Application US/10453372  
Publication No. US20060003323A1  
GENERAL INFORMATION:  
APPLICANT: Alsobrook, et al.  
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
FILE REFERENCE: 21402-589 A  
CURRENT APPLICATION NUMBER: US/10/453, 372  
CURRENT FILING DATE: 2003-06-03  
PRIOR APPLICATION NUMBER: 09/789390  
PRIOR FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: 60/185967  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 09/823187  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/195792  
PRIOR FILING DATE: 2000-03-10  
PRIOR APPLICATION NUMBER: 09/839446  
PRIOR FILING DATE: 2001-03-19  
PRIOR APPLICATION NUMBER: 60/199476  
PRIOR FILING DATE: 2000-03-25  
PRIOR APPLICATION NUMBER: 09/863776  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: 60/208263  
PRIOR FILING DATE: 2000-05-31  
PRIOR APPLICATION NUMBER: 09/939398  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: 60/227800  
PRIOR FILING DATE: 2000-08-25  
Remainder prior application data removed - See file wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1609  
SOFTWARE: Curaseq1ist version 0.1  
SEQ ID NO 116

```

; LENGTH: 2765
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-116
Query Match
Best local similarity 19.5%; Pred. No. 3e-09;
Matches 365; Conservative 251; Mismatches 652; Indels 603; Gaps 99;
3.1%; Score 272.5; DB 6; Length 2765;

```

49 GHIVANG-NIGPLPLTLSTPLNKTDIGFGIFNGLSYDRKNSLSLSTGBNTK--- 104  
Db 1098 GHIFOKSFQASPMIASTFIW--DKTDA-----YGORVYGLSDVAVSV--GFEYETCP 1145  
Qy 105 --VLETDKYTKLOQKLD--NL--RPEK-----DLKENCRIIHK--SGDIEVLTG----- 147  
Db 1146 SLTLMKERTALLOGFELDSPNLGMSLDRKHLILNVSQ--LHKGGENQFLIQQPAIL 1202  
Qy 148 -----FNNAFPLKVP-----KKLNPAGHAIYIDWNEATQPLRLRIYDDLGDHDP 195  
Db 1203 TSIMNGRRRSISCPSCNGLABGNKLLAPVALVAGIDSL----- 1242  
Qy 196 LNLBYOGLIKITLTLFPGQKSGYRTBLRFLNRQL-----NSINFSLGNENPLTWSPGYT 251  
Db 1243 -----YGDGFNYIRRIFFPSRNTVSIIELR--NKEFHSNNPAKHXYLA--VDPVSGSLYVS 1294  
Qy 252 PIKNGILQWITSMTAPGLKETVYVSNNOGHHFPQSANLVLPLYTLMKQVPGAGP 311  
Db 1295 DTNSRRITV-----RVKSLSGTKDLAAGSEVVAQ-----TGECLP--DEARCGDGK 1340  
Qy 312 AIQAEYSYSHNVYGGSGNGINW-----NKLDNLVGLMTEVNYGSGTESRRYK---DK 360  
Db 1341 AIDA--TMSPRGIAVDKNGMLFVDATMIRKQD--NGIISTL--LSGNDLTARPLSCDS 1396  
Qy 361 EGHDIQVRIE-----RTYNNYHLTSECKQONGYIQTETAYAIIGHNFDOSQSF 412  
Db 1397 SMDVAGVRLWEPDVLAVNPMWNSLYVL-----ENNYILRTTENHQVSIIA-----GRPMHC 1447  
Qy 413 QLPKTEETWRKADNSYRSSEITETFDSEGNPLTKYIKKKYKIKIISPTHMEYPPAGE 472  
Db 1448 QVPGIDYSLSKLAHLSBASALAI SHGVLYITETDEKIRLRLQVTTNGEICLLAGA 1507  
Qy 473 VDNCPPE-----PYGFRFVKKIIQTPYDSEFPDOP 503  
Db 1508 ASDCDCKDVNVCYSGDDAYATDALNSPSLAVAPDG--TIYIADLGNIRIAVSNKR 1566  
Qy 504 --EKFIQYRSLIGSQSHVTLKIEBHYSATQULNSTL---FOYNTDKSELGLKQTEC 558  
Db 1567 VLNAFQYEAASPGBELVFNADGIHQYVSLVGTGEVLNFTYSTD--NDVTELI----- 1620  
Qy 559 TKENGKITSVNHKFTYTKODDTLQOSH-----SITHDNFTIHRQ----- 600  
Db 1621 --DNNGNSLKI-----RRDSGMPRHLLMPDNOIITLVGNGGLKVVSTNLEBGLM 1671  
Qy 601 VRSRYTGRLEPSDITDVIQWMSYDKLGRLLTRTLNSGT-----PYANTLVYDVELNLT 654  
Db 1672 TYGNTGLATKSDDEGTFTFYDHBGRLLTNVTRPTGVVTSIHKREKESITDIENSNR 1731  
Qy 655 QDDNRPEFVITTTDVNG-----NOLNNEPD-----GAGRH----- 684  
Db 1732 DDD-----VTVINLSVSEASVTVQDVVNSYQLCNGNGLRVMYVANGMISHSSEPHVL 1786  
Qy 685 -----VSQC-----LKSDSDGKFTYTHHQYQEGHHHTSYSDYL 721  
Db 1787 AGTITPTIGKCNISLPMENGLNSIEWRLRKQIKG--VIFGKRLRVHGNLSI--DYD 1843  
Qy 722 TNGR--OOTPDKTHLSMSKSYDNWGO-----IANTHWSY-----GVSEK 759  
Db 1844 RNIRTEKIYDHRKFTLRIRIYDQVGRPLMLBSSGLAAVAVSIFPRGLAGLQGMASER 1903  
Qy 760 ITVDPTTLATYKQLQGSNNVNOTGKEVTTYTSPSQPIQITLTFDEAGHLSCHTLTRDGWD 819  
Db 1904 TDID-----KGRIVSRMFADGK--VMSYSYLDKSMVLLLSQGRQY----- 1942



```

QY 820 RVRKETDAIGQCTIYOQYDNVNRVIOITLPDGTIVNRKAPFSTDTLLITDIRVNGISIGQQ 879
Db 1943 -----IPEYDSSRLLAVTMPB---VARR--SMSTHTSIGYR---NINP 1980
QY 880 TFDGLSRLTOSOGGRVWAAVYYSAGNDCEPSTVITPDGQFIHQY--OPELIDAVLOVAS 937
Db 1981 PESNASVIFPYSDDGRIL-----KTSFLGTGRQVFYKYGKLSKLSSEIYVD--S 2026
QY 938 NEIQOQSPYNPVGAL--LKAVAEGQSLTPIYPSGRLKMNENIDMKKMSYLWTLRGLN 995
Db 2027 TAVT--FGYDETTGLKKNVNIQSGFSCITIRKIGPL-----VDQITRESEBGMVN 2077
QY 996 GYTDLTGTIOKISHDTHGRVTOIKDSSIKTTL-----NYDDL--RHIGS-----QVT 1041
Db 2078 ARFPDY-----YHDSFRILASIKPVISETPLPVDIVRYDEISGKVEHFGKGVYIYDIN 2131
QY 1042 DLATGMLTTTVEPD--GLNREIGRKLCDSSGHTLIDIQOSWLKTOQLANRIVKLN---GV 1096
Db 2132 QIITTAVMTLSKHFDTHGRIKEVOYEMFRSILMMWTVOYDSM-----GRVIRKELKIGP 2185
QY 1097 LQRTQOYSYDSRNLNQKCGA-----ECPTDKYGHSLVTQNFYDIQNTTACHTT 1149
Db 2186 YANTTKYTYD-----YDGGQLOSAVANDRPTWRY-----SYDLGNL----- 2223
QY 1150 FADGTEBHATPKFANPTDPCQLTVEHHTHPMPDNIRLKYDKAGRVINI-----TDNHG 1203
Db 2224 -----HLNPGNSVVL-----MP-----LRVLDLRIRITRLGDOYQKIDDDG 2259
QY 1204 -----NTENPTYDTLGRIL-----ONGQGSVYGVDPLNRLVSQKTDPLDCELYRETMLV 1252
Db 2260 YLCORGSIDIFEYNKSGILLTRAYNKASGWSVOYRYDGVRRASVYKTN--LGHNLQYFYSDLH 2318
QY 1253 NEVR-----NEMIRLL--RTGETIIAQORASKVLLTGTDSQGS--VLTSDKQNL 1300
Db 2319 NPTRIITHVYNSNSEITSLYYDLOHLFAMESSGEERYVASDNIGTPLAIVSINGLMIX 2378
QY 1301 QEAYSAYGK--HKSTANDASIIIGYNGERADPVSGVTHLGNGRSYDPTLMRFHTPD--- 1354
Db 2379 QLOYTAVGEIYYDSNPDQFWIGFPHGLYDPLTKLVHFTQ--ROYDVLAGRMTSPDYTMW 2436
QY 1355 -----SLSPFGAGINPYSCLGDPINRS-DPSGHL-----SWQAWTGI----- 1392
Db 2437 KNVGEKEPAPF-----NLTFMFKSNPNLSSELDLKNYVTDVKSWLVWFGFOLSNIIIPGPRA 2491
QY 1393 -----GMGIAGLLTIIATGMAIAAG-----IAAIA----- 1421
Db 2492 KMIVVPPPEYELSESQASENGQILITGVQOTTERHNOAFMALEGQVITTKLHASIREKAGHW 2551
QY 1422 -STSTTALAFGAL-----SVTSDITSIVSGALBDASPASISIG----- 1459
Db 2552 FATTTPIIGKIMFAIKGRTYGVSSIAS--ED-SRKVASVLMNAVYLDKMHYSIEGK 2607
QY 1460 ---WVSKMGAGAAGIA-----ESA-----IKGCTKLATH/GAFAEDGE 1493
Db 2608 DTHYFVKIGSADGLVLTGTITIGRKVLSEGVNVTVSQPTLLVNGTRBRTNI---EPOY 2663
QY 1494 NALKSTSESSEIRIKWGV--RSIDREIVNNEGQVTKHRSRGYTDNFMKGEOALIVHGD 1552
Db 2664 STLIS-----IRYGLTPDTLDEKAR-----VLDAQROPALGTAMAKOQKARDG-R 2710
QY 1553 DGFLYHTEGNK 1563
Db 2711 EGSRLMTEGK 2721

```

Search completed: January 30, 2006, 09:53:34  
 Job time : 29.3373 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2006, 09:41:43 ; Search time 29.1754 Seconds  
(without alignments)  
5517.344 Million cell updates/sec

Title: US-09-889-874A-23

Sequence: 1 VYIKFKLRRRTMSDNNE.....PKIIIGRTKTKPKTPRP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 263416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 263416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: PIR 80:\*

1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1214	13.7	709	2 S38241	hypothetical prote
2	912.5	10.3	528	2 S38242	hypothetical prote
3	432	4.9	2334	2 S32920	cell wall-associat
4	429.5	4.8	336	2 S38239	hypothetical prote
5	385	4.3	2167	2 APL489	cell wall-associat
6	381	4.3	2183	2 T37218	hypothetical prote
7	365	4.1	1400	2 E90886	RhsB core protein
8	353.5	4.0	843	2 AB0539	RhsE-family protei
9	342	3.9	1397	2 C64805	RhsC protein precu
10	341.5	3.8	1404	2 E90658	RhsG core protein
11	341	3.8	1394	2 H91236	RhsH core protein
12	340.5	3.8	1411	2 E65155	RhsB protein precu
13	339	3.8	1426	2 H64780	RhsD protein precu
14	336.5	3.8	1404	2 E85509	hypothetical prote
15	331.5	3.7	1377	2 C65159	RhsA protein precu
16	331	3.7	1409	2 F91187	RhsA core protein
17	326	3.7	1399	2 A99720	RhsC core protein
18	322.5	3.6	1377	2 E86034	RhsA protein in rh
19	319	3.6	1397	2 A85570	RhsC protein in rh
20	316	3.6	1398	2 E85549	hypothetical prote
21	313	3.5	1398	2 H90698	RhsD core protein
22	297	3.3	1512	2 AH0439	probable membrane
23	280	3.2	1317	2 F83310	conserved hypochet
24	278	3.1	985	2 B86084	hypothetical prote
25	276.5	3.1	1438	2 A10093	conserved hypochet
26	271.5	3.1	1354	2 AG0538	Rhs-family protei
27	266.5	3.0	2515	2 S47008	tenascin-like prote
28	248.5	2.8	1616	2 E90704	Rhs core protein w
29	245.5	2.8	1645	2 H85554	hypothetical prote

30	245	2.8	2825	2 T14271	Dcc4 protein, stre
31	230	2.6	794	2 T36972	probable membrane
32	230	2.6	849	2 T46253	hypothetical prote
33	220.5	2.5	1185	2 A42404	collagen adhesin -
34	219.5	2.5	1962	2 A32634	lactocerin (EC 3.4
35	218.5	2.5	2406	2 A54148	odx protein - fru
36	217	2.4	356	2 T37136	hypothetical prote
37	216.5	2.4	2894	2 C64474	hypothetical prote
38	215.5	2.4	2044	2 AB1180	probable peptidogl
39	213	2.4	1487	2 AG2560	hypothetical prote
40	211	2.4	3283	2 AC1018	large repetitive p
41	208	2.3	2703	2 H81193	hemagglutinin/hemo
42	207	2.3	2893	2 A64556	coxin-like outer m
43	205	2.3	1902	2 S05997	lactocerin (EC 3.4
44	203.5	2.3	1366	2 S57664	IGA-specific metal
45	202.5	2.3	1349	2 A11476	cell surface prote

ALIGNMENTS

RESULT 1									
S38241									
hypothetical protein - Coccidia burnetii									
C:Species: Coccidia burnetii									
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004									
C:Accession: S38241									
R:Title: D.; Willem, H.; Haas, M.; Krause, H.									
submitted to the EMBL Data Library, October 1993									
A:Reference number: S38215									
A:Accession: S38241									
A:Status: preliminary									
A:Molecule type: DNA									
A:Residues: 1-709 <THI>									
A:Cross-references: UNIPROT:Q45948; UNIPARC:UPI0000BA5E9; EMBL:X75356; NID:9407370; PIR:									
Query Match 13.7%; Score 1214; DB 2; Length 709;									
Best Local Similarity 40.8%; Pred. No. 2e-59;									
Matches 287; Conservative 101; Mismatches 280; Indels 36; Gaps 16;									
QY	554	KQTECTKGENKGYTVVHKFTYTKODDTLQOSSHITTHDNFTIHSQVRSRYGRLEFSDT	613						
DB	3	RRAEVLTSKKEGKKXQNGTTFPAISSQAHLQKIDFTGSDGKTKTSISEGRYSGRLLST	62						
QY	614	DTKDIYQMSYDKLGRLLRTLN-SCTPYANTLTYYELANLQDDNRPPVITTDVANGN	672						
DB	63	DEIGNVYQYDEIGRLITQTVNASTTYASTTYSLETDRGKVTAKITVTTDPKGN	122						
QY	673	QLNEPFGAGRAHVSQCLKSDG----DGKFTYIHQYDEGRHHTSTSDYLTNGRQ	727						
DB	123	QLRTYDGLGRNLKQERLDDAAVSOQTGTWYITTHQOQDALGRSKITTDVLT	177						
QY	728	TDDPKVAH---LSMSK--SYDNMGQIANTHMSYGVSEKRTVDPTLTATKQLOSNMNVQ	781						
DB	178	LDBEVGHAGSVLSTSKYMHDSWGQNLTVFSQGYGRSYVDPTTRAT--LDPSGSQK	235						
QY	782	TGKAVTTTPSSQPIQITLFDKAGHLQSGHTLRGDRVRKSTDAIGCCTIYOYDNYNR	841						
DB	236	LGQQLTEYNLAGPIKITYQDSQTEGSAHYEDGIGQLRKEDLGOITLYEYDHFGR	295						
QY	842	VIGITLPDGTIVRKAPSTDLTDIRVNGISLGQGTGSLRLOSDGGGVAVYTY	901						
DB	296	VITQTLPENTIIQISVAPHSITASLTGISVNNFSGMGQITDSLERLTETISGGTSPSY	355						
QY	902	SAGNDQPSVTITPDGQFIHYQYQPELDDAVLQVANSNITQPSYNTGALLKAV-AEG	960						
DB	356	ENMS-SVPAAVTAPVGTSTVYELKELGNAVKKISABEILQTMVDYDLTAMNSATQAAG	414						
QY	961	QSLTPIYPSGRLLME-NIND---MKMSYLMTLRGLENQYTDLTGTIQKISRDTHRVY	1016						
DB	415	MIRQMTYPSGLKNKETSMPDGAOKSTAYTAYLAGAPOSTYVFGVQRYDYDDEHRR	474						
QY	1017	QINDSSIKTTLNVDLNRHIGSQVTDLATGMLTTTVEPGLNREIGRKLCDSGHTLDI	1076						

```

Db      475 GIBNDIKVSLDYDAFGRFTKQATDKKTGAIVLSTTLTYDDLNREIKREISASQSVLVI 534
Qy      1077 QQSMLKTOQLANRYKANGVLTQRTBOYSYDSRNRLNQKCGACFPDKYGHSTVQNT 1136
Db      535 EGYTORHNLKERITTORRTTLRKEMFAYDSRNRLIEYTCGEARPODYPKALHRTQFS 594
Qy      1137 YDIIGNITACHTPADGTEHDATPEFANPTPCOLTVEYHTHPMPDPNIRLKYDKAGRVI 1196
Db      595 YDALGNMTKOTDPSGG-RMTATYIY-SADIPTOCLKNNHSDPKREITTEYDKAGMT 652
Qy      1197 NITDNHNTENFTYDTLGRLO-NG--QGSVGYDPLNRLVSQ 1235
Db      653 R-DEAGRT--LRYDALGRLOQVNGAGAKGGQYADVLTNLVSQ 692

```

## RESULT 2

```

S38242
hypothetical protein - Coxiiella burnetii
C/Species: Coxiiella burnetii
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C/Accession: S38242
R/Title: D.; Willems, H.; Haas, M.; Krauss, H.
submitted to the EMBL Data Library, October 1993
A/Accession: S38242
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-528 <THI>
A/Cross-references: UNIPROT:Q45949; UNIPARC:UP100000B0C0B; EMBL:X75356; NID:g407370; PIR
C/Superfamily: Coxiiella burnetii hypothetical protein

```

```

Query Match 10.3%; Score 912.5; DB 2; Length 528;
      Local Similarity 40.8%; Pred. No. 6.6e-43;
Matches 222; Conservative 72; Mismatches 203; Indels 47; Gaps 17;

Qy      18 NEF-FTQANFTSAVSGGVDPRTGTYNIQITLGHVNGNLGPTPLTLTSLSPINKTDI- 75
Db      2 NELPYQTQNTFISAVQGGVDPRTGTYNIMVLAELTGNDNLGPDLPLFLANTSHLSTNIC 61
Qy      76 GFGIGFNGLSVYDRKNSLSLSTGENYKVIETDKTVLQOKKDNLRFEDLKENCYRI 135
Db      62 GFGIGCSVGISIDYKNNMLLISGGERKTYDMDSDVYRQKKNPFER--IKNGYII 119
Qy      136 IHKSGDIEVLTG--FNNAFPLKVPKCLNPAHAIYIDMVFBEATQPLNRYDLDLDDHD 193
Db      120 KYKNGKTYLNYKKGDLF--LPQKIFSPLGMPKLKTWENRQYVNLTKIEDAKD--- 173
Qy      194 IPLNLLEYQGLKITLLTFPGQKEGYRTEFLRNQLNSINFLSGNENPLTWSFGYTP 253
Db      174 -VLCIKIDYQFSDMARITFWPKTESYQDLDFVNEYLWVTKNSRSR--LWSPFNVDV 230
Qy      254 GKNGILGMITSMAPGSLKKTETVYNNQGHHPQSANLVLPVYTLMKQVPPAGQPAI 313
Db      231 GAGNFT--LTVQKSPGLTETVYQAGV--RPDESCKALPSVYNYRQSPGKGPDI 285
Qy      314 QAESYTSHTNYVGGSN--GIWNKLDNLVG-LMTEVYVSGTESRRYDKCGHDQIYVIE 370
Db      286 VKETETVSNLTGICASLGKAMNEDENIYVMDVYISTEKLIVNRE---LVSIS 341
Qy      371 RTYNNYHLTSECKQONGYIOTETAYVAIIGHNFDSPQSPQLPKTKTETWR-SADNSY 429
Db      342 RIYNSYLLISETTRQNSCEVIVETDYAKPGLSPDKQPKQPLPKETKMTWRENSKQC 401
Qy      430 RSEITETTFDESQNPULTVYIKDKTKQKILSPSTHMEYTPPAGVDN---CPPEYGTTR 485
Db      402 RSEITTTTFDEGNLLTKIEPD-----GRTETIYVDSKGETDKGIVLCPPEPNQVVR 454
Qy      486 FVKKIIQTPYSEF-----KDDPEKIQYRSLISGSHVTLKIEBRHYSATQQLNSTLFG 541
Db      455 FVKIQIYVPADESEFAVYQQTYYAYAOYPCIASSLSIAVLTQIETLCSDDL---LIT 510
Qy      542 YNTD 545

```

```

Db      511 INTD 514

```

## RESULT 3

```

S32920
cell wall-associated protein precursor wapa [similarity] - Bacillus subtilis
C/Species: Bacillus subtilis
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C/Accession: S32920; E69730; T47101
R/Title: Mol. Microbiol. 8, 299-310, 1993
A/Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis
protein.
A/Reference number: S32919; MUID:93302506; PMID:8316082
A/Accession: S32920
A/Molecule type: DNA
A/Residues: 1-2334 <EOS>
A/Cross-references: UNIPROT:Q07833; UNIPARC:UP10000060C90; GB:L05634; NID:g304177; PIDN:
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertek
C.; Bron, S.; Brouillet, S.; Brunsch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fultz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gallier
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A/Authors: Schlecht, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowka, A.; Serot,
akeuchi, M.; Yamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Wintere, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A/Authors: Yoshikawa, H.F.; Zunsstein, E.; Yoshikawa, H.; Danchin, A.
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: E69730
A/Status: nucleic acid sequence not shown; translation not shown
A/Accession: S32920
A/Molecule type: DNA
A/Residues: 1-2334 <KUN>
A/Cross-references: UNIPARC:UP10000060C90; GB:299124; GB:AL009126; NID:g2636442; PIDN:CAF
A/Experimental source: strain 168
R/Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
Microbiology 141, 337-343, 1995
A/Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome conta
A/Reference number: 224350; MUID:95219088; PMID:7704263
A/Accession: T47101
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2334 <YOS>
A/Cross-references: UNIPARC:UP10000060C90; EMBL:D31856; NID:g603765; PIDN:BAA06656.1; PI
A/Experimental source: strain BGSC1A1
C/Genetics:
A/Gene: wapa, N176
C/Superfamily: cell wall-associated protein wapa

```

```

Query Match 4.9%; Score 432; DB 2; Length 2334;
      Local Similarity 20.4%; Pred. No. 4.1e-15;
Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;

```

```

Qy      44 IQTLGHIVG-NGNL-----GPTPLTLTSLSPINKTDIGFGIGNFGLSVYDR 90
Db      896 IDIPSGQLNATGNGVINNEBDLSIDGRPGIGLSRTYNSLSDSDHLFGQGW-----YADA 950
Qy      91 KNSLSTSGENYVITEIDTVLQOKKLDN-----LRFEDLKENCYRIHKSQDIEVL 145
Db      951 ETSVISTDGGAMY--IDEDATHTRTKADGTYQPPGVVLELTADQFLTKTD----- 1004
Qy      146 TGFNNAFPLDKVPKCL-----NPAGAIYIDN-----FEATQPLNRIYDLDGH-- 192
Db      1005 ---QTNAVFKKGGKGLQKVVDDGNHNAIVTYINDGNQLTATDASGRKLTFTYDE-NGHVT 1060
Qy      193 -----DIPNLLEY--QGLI-----KTIITLVTP 214

```

```

Db 1061 SITGPKKAVTASYENDLLKKVTDIDGTVTSYDSEGRVLKQVANSSTAEKPFVTEY-- 1118
Qy 215 QKGGYFELAFRLQRLNSIHNF-----LGNENPLTMSFGTTPIGKNIILQOMITSMF 267
Db 1119 QSGHRLREKAKINAKKETVYVSADAKKTLTMTQPNGRKVOGYNEANP-----IQVID 1172
Qy 268 APGLKETVA--YNNN-----QGHFPOSANLPVLPLYTLMKQVAG 309
Db 1173 DAGGLKTTTKTEGNNVEDVDNDVGTGKATESYDQGN-----VTSYKDAYGT- 1225
Qy 310 QPAIQAESYTSYSHNYVGGSGNIMNNLNLGYLMTENYSGTESRRYKQKEGHOIVRI 369
Db 1226 -----ETFEYKNNNDV-----TKMKDEGNVTDIAYDGLDAVBETDQSGSSAAV 1271
Qy 370 ERTYNNYHLLTSECKOONGYIQTETAYVAILIGNFPQSOPQOLPRTKTEYRSADNSY 429
Db 1272 YDKRGNIQSSKSLASNTILK-----DGSFEAQKSGNNLTASKD----- 1311
Qy 430 RSEITETTPDESNGPLFKVILKDKTKIKIISPS-----THMEYPPAGEVNDCCPEPYGPT 485
Db 1312 RKKIS-VIADKSG-----VLSGSKALEVLSOSTSAGTDHGSATQVLEPNTTTLG 1365
Qy 486 FVKKI---QTPYSEKDDPEKFIQY---RYSLIGSOHVTLKIBERHYSATQLNS-- 537
Db 1366 KITDLAKSAYFNIIDLRDQKRIQWIHNEYSALAKNDWT-----KQITFTTPAAGK 1421
Qy 538 ---TLQYNTDKSELGL-LKQTECTGKNGKTYSVHAKTYTKODDTLOOS-HSITTHD 592
Db 1422 AVYMEVDHDKDKGKAMPDEVOLEKEGVSSSNPQNSFTSATENMNVSGASVDSSE 1481
Qy 593 NPTIHSQVRSRYTGLFSDTDKDIQMSYDGLRLTRT-----L 635
Db 1482 GFNDVSLKAKRTSASQASVTKQTVLQSGANDKPVYLLTLTGSKASVYFTDEKDYSL 1541
Qy 636 NSGTPYANLTYDYEL---NNLQDNRPPFYI----- 664
Db 1542 QANVTAYADSGTGYNAKFPSTQEMNAAVVIPTKPIKINKVDLSILFQKSAITGVWDDI 1601
Qy 665 -----TTDVGNGQLRNEPDANGRIVSQCLOSDGKRTYTHTOQYDEQGNHT 714
Db 1602 RLIEGSLITKSTYDSNNGYVTKEDDELGYATS---TYDETGK---KTSETAKGKKT 1654
Qy 715 STY--SPYLTNGRQQTQPDKVHLSMSKSYDNWG-QIANT-----HMSYGVSEKI-- 760
Db 1655 YTTDOAQQLTNMTLSNGTSIHH---SYDEGNEVSKTIRAGADQTYKEBYDMGLVAK 1709
Qy 761 TVDPITLTATKOLQSNNSNVQT---GKEVYTYTPSOQPIQITLFDRAHQLQSCHTLTRD 816
Db 1710 TTDPDLGNVLASEVDANGLTKTISPNGNV-----SLSYD 1744
Qy 817 GMDRVRKETDAIGCCTIYQYDNTRVYQITLPDGTIVNRKAYPSTDTLITDIRVNGISL 876
Db 1745 GTRVKSYSYNGEYKIFTYDKNGN-----ETSVNKEQN-----TT 1781
Qy 877 GQGFQDLSLTQSDGGRWAYTYASAGNOCSTVITPDGQFIHYQYQELDAVLQVA 936
Db 1782 KKRFTPKKRNRLTETLDRGSGQWTWYPSDCLKTF---SWHH-----G 1821
Qy 937 SNETIQQFSYNPVYGLLKAVAGOSLTPYYPGRLMENINIMKMSYMLTLRGLENG 996
Db 1822 DQKGTGNFTYN-----KLDQMIEKOSTSYSSTDYEN- 1854
Qy 997 YTDLTGTIQKISRDTGRVTOIKDSSIKTTLNVDLNRHIGSQVTDLATGMLTTVFEFD 1056
Db 1855 -----GNVQ-----TFITNGCGTSSFSYDERNLVLSIHIGDKNGGDIILTESYEV- 1898
Qy 1057 GUNREIRKLCDSGCHTLDIQOSWLKTOQLANRIVKNGVLRQREOYSYDSRNLNLYKC 1116
Db 1899 -----DAANGNTTINS-----ASGVK---QYBYGKLNQLVK-- 1927
Qy 1117 DGAECPDYKXGHSIVTQNFYDIYGNITACHTTEADGTED--HATFFRANPTDCCQLETV 1174
Db 1928 -----ETHHDEGTIV---EYTYDGFGRKRTV-TTKDSSSKTVANSFNIMN-----QLTLKV 1973

```

```

Qy 1175 H-----HTHP-DMPDN-----RLKYDKAGRVINITDNHG 1203
Db 1974 NDESISYDKNKGRNTSGKTYTTWDAEDNLTAIVYKKGEDKPRATYKIDDEKGNRIQKTVN-G 2032
Qy 1204 NTEPFYDTLGRLONGQSGSVYGDPLNRVLSOKTDTLDELY-YRETMVNEVRNGEMIR 1262
Db 2033 KVTNIFYDG-----DSLNVLYETDADNNVTKSYTGD-----SGQLLS 2070
Qy 1263 LMTGETIIAQCRASKVLTLGTDSSQSVLITSKQULSQEAYAYGK-HKSTANDA---S 1318
Db 2071 YTEGKKYFYHVYNAHDIIAISDTGKTV-----AKQYDAMWAGPTKTESDEVKDN 2122
Qy 1319 ILGNGERADPVSGVTHLNGYRSYDPTLMRFHT--PDSLSPGAGINPYCYLADPPIN 1376
Db 2123 RRYAGYQYDEETGLYLMARY--YEPKNGVFLSLPDPESDSDSLDQNGYAGNNPVM 2180
Qy 1377 RSDPSGHLISQWATGIGMGIALLLTATGMAIAAG---GIAAIASTSTTALAFA 1432
Db 2181 NVDPDGH--W-VMLVYNAGFA-----AYDGYKAYKSGKGMKAAMAAASNPFGPKIFG 2231
Qy 1433 LSVTSDITSIVSGALDASFKASSIIGWVSMG---GAAGLAESAIKGTATLHAGF 1488
Db 2232 ASRAYFTK-----KAVKITGHTHGHGLOSIGRNG-----GRGVNLRADKANA- 2273
Qy 1489 AEDGENALAKSTSESRIRKMGVTRSLDRE--IYRNEGOYIKDHSRGYTDNFMKGEGAL 1546
Db 2274 -----VRSPPKVTIKQPNAGTKYVGCATVVLNRGKVTIAYG-----SS 2312
Qy 1547 LVHGDKGFLYHTEGK 1563
Db 2313 RAKSGKHVFHTHGK 2329

```

```

RESULT 4
S38239
hypothetical protein - Coxiella burnetii
C:Species: Coxiella burnetii
C>Date: 20-Feb-1995 #sequence, revision 20-Feb-1995 #text_change 09-Jul-2004
C/Accession: S38239
R:Thiele, D.; Williams, H.; Haas, M.; Krause, H.
submitted to the EMBL Data Library, October 1993
A:Reference number: S38215
A:Accession: S38239
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <THI>
A:Cross-references: UNIPROT:Q45946; UNIPARC:UPI0000086171; EMBL:X75356; NID:G407370; PDB

```

```

Query Match 4.8%; Score 429.5; DB 2; Length 336;
Best Local Similarity 34.3%; Pred. No. 26-16;
Matches 121; Conservative 48; Mismatches 91; Indels 93; Gaps 13;

Qy 1348 MRFTPDLSLSPFAGGINPYCYLADPPINRSDPSGHLISQWATGIGMGIALLLTATG 1407
Db 1 MRINCDSWSPFAGGINPFAICDGPINRVNDNGHLSWQAEIDIGLVGVLAIVPTAG 60
Qy 1408 MAIAAGIAAIASTTALAFAALSVTSDITSIVSGALEDASPKXASITLGVNSMGMA 1467
Db 61 TSAAGAIASALIASAIAISLVVGTGVAADVASIASGALEDANPQASATLGWISLGLG 120
Qy 1468 AGLAE--SAIKGTGLATLHAGFAEDG-----ENALLKSTSESRIRKMGVTR--RS 1513
Db 121 PGAVSGLATTAARAGKGL--ISGLAKGGKIRSQSPVQGISYSLSGDPLRGGPFPFOS 177
Qy 1514 LDREIVANEE-----GOVTKDHSRGYTDNF-----MGL 1541
Db 178 LSRVYTAPEBMRPAGLNYMHKVSQKSLGQYHFGADRLEIFGYEIRPEIFRRRPSITK 237
Qy 1542 GEGAIL-----VHGDKGFLYHTEG-----NKNNGKGPYTRHPTPEQLVDY 1561
Db 238 RDIIVILSGTHGRVADN-----WTSQGLRRPDILERAIFYIEDVQNYKG-----QLNGR 285

```

Query 1582 LKDNINVDLTCGG-----DKPVLHLSCKYKSSGADKKAKYIN-RPVAY 1625  
Db 286 VKIYDAMGMITSESEGRVAVRNQHVILIGYCKGRNDEA---LTYHRILRPVISTY 335  
RESULT 5  
AF1489  
cell wall-associated protein precursor wapa (B. subtilis) homolog lin0454 [imported] - I  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C;Accession: AF1489  
R;Glaser, P.; Farnagel, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseigneur, O.; Entlian, K.D.; Fahl, H.  
D.; Jones, L.M.; Karsel, U.  
Science 294, 849-852, 2001  
A;Authors: Krefel, J.; Kuhn, M.; Kuntz, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueder, T.; Simoes, N.; Tiberet, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland,  
A.;Title: Comparative genomes of Listeria species.  
A;Reference number: UNIPROT:Q92EK5; UNIPARC:UP100000CC234; GB:AL592022; PIDN:CAC95686.1;  
A;Accession: AF1489  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2167 <GLA>  
A;Cross-references: UNIPROT:Q92EK5; UNIPARC:UP100000CC234; GB:AL592022; PIDN:CAC95686.1;  
A;Experimental source: strain Clijp1262  
C;Genetics:  
A;Gene: lin0454  
Query Match 4.3%; Score 386; DB 2; Length 2167;  
Best Local Similarity 19.4%; Pred. No. 1.3e-12;  
Matches 327; Conservative 212; Mismatches 582; Indels 568; Gaps 67;  
Query 17 NNEFTQANNFTSAVSGVDPRGLY---NIQITIGHI-VGNGNL-----GP 59  
Db 718 NNTTSVSEPRTPSTPEALNKQLGKVDYWTSPVRGSEVNAATNGFLFHEHEDPMLBGRGP 777  
Query 60 TLPLTSLSPKNTDIDGIFGIFNGLSVYDRKNSLSLSTGENYKVLETDTV----- 112  
Db 778 SINNRFTNSQDDATGIFGKMTSTL-----EKLVEENG-NLWVVSDDKIHRTKGG 831  
Query 113 -----KLQOKLD-----NLRFKDLKENCYRIIHKSGDIEV 144  
Db 832 DKYAPPGIYSEITKMAADGVLKIEEDKSETRFLVDGRKSEKDTGKMLTYEYTDGKLTG 891  
Query 145 LTGNNNNAFDL---KVPKILNPAGHAIYIDMNPENATQPLNRIYDLDHDDPLNLTLE 200  
Db 892 LRDSAGRTVTLTYEGELVKELVGPEDRKISYTN-----D 926  
Query 201 YQGLIKTILTFPGQKEGYRTLEPLNQNSIHNFSLGNEPPLTWSFGYTPIGKNGILG 260  
Db 927 KQELISSSTARGLKYRYGTYDGL-----LTSIYDPKHTBEKPYETTPAYEE----- 972  
Query 261 QMITSMTAPGLKETVNSNNQG-----HHFPOSANLPVLRYTTLAKQVUGA 308  
Db 973 EKLTEITDPVKKKTTLSDYKAEQQTTLTNEKKKTIYSYNDAGN-----PKKEIYDA 1024  
Query 309 GQPAIQAEYSTSHYVGGSGNGINNNKLDNLGLMTEYNGSTESRRYKXKEG----- 363  
Db 1025 DGLKLTITTYESSN-----LVKENVKQGEETAYADAGNITAT 1065  
Query 364 DOIVALERTYNNMHLTSECKQONGYIQTETAYVAIIGHNFDQSPQFOLPKTETETWR 423  
Db 1066 DAYGRESTTYNDNNVTS-----STDTE-----GRKTTVYD 1097  
Query 424 SADSNSYSEITE-----TTPDESGNPL-----TKYIK 450  
Db 1098 GADAVSETLATESQVSVYQDAYGNPIRGSELSSGNLQNSGFEKAGAVSNMTLLQS 1157  
Query 451 DKK-----TQ-----KISP-STHMBYYPAGEVYVNCPEPYGFTFRFYK 489  
Db 1158 DAKSGMTDNTQSAFGALGGSGSVKLTSEANSIVKGYSSVYORVAVEETTYTTFAMAKT 1217  
Query 490 -----IIOPTDSEFKDPEKFIQYRYSLSIGSOSHVTLKIEER-HYSATOLLNSTL 539

Db 1218 SGMTNADALLIGRLQDANAKVDYDAGVW-----QSNPATSIKKQGDVAVKQL----- 1264  
Query 540 FQNTDKSELGRLL-----KQTECTGENKKTYSVYKFTYTKQDDT 581  
Db 1265 -TPTKSNTRQVLLYLDNEQDPAHKKGTIWDNVQPEKGSVASSINPVNNSFEENHGT 1323  
Query 582 LQOSSHSTTHDNFT---IHRSQVRSRYRGLFSDTDTKQIVTQMSYD-----KLGRLLTR 633  
Db 1324 LPTGMWRITGATLQAKVVDNNGSHSGDSAVYFRKATSEAVYTHIVQDVPVNGEAKALYT 1383  
Query 634 TLNSGTPYAN-----TLTYDEL-----NNLQ-----DNRPPVITTTD 668  
Db 1384 SALSKEADAKANGSAVATMSNDYSVWGTVVYQDGTSTSSQCPPLGTNDMNRSAVVAVPTK 1443  
Query 669 -----VNGNQL-RNRPDGAGRHVSQGLKQSDGDKREYTI 701  
Db 1444 PVKAIKYVTMFRNGLTGKAMPDVRVYEGVLTNREYDASGNV----- 1487  
Query 702 HFOQYDEGHRHTSTYSDYLTNGRQOQTPDQKHLMSKSYDNMQIANTHMSYGVSEKIT 761  
Db 1488 -TASYDEGRKISPTDYI---GNKTSFEDKGNKKTLYTDADNALIDTLANGSTVAVK 1543  
Query 762 VDPITLTATQOLQSNNSNVOTGKEVTTTPSQPFI-----QITLPEAGHLQSCHTLT 814  
Db 1544 YD-----DNGNTTEKNVTASGKTOKRIYEYDVNDKITAFTDA-----LNRTI 1585  
Query 815 RDGMDRVREKTDAL---GQCTIYQYDYNVRVIOITLPDGTIVNRKVPFSDTLITDIRV 871  
Db 1586 KYREDAAGNETKAKIMPNRGVTESTYDSADMDGIMKADKAFKQYDPNNGQTYVTD-EI 1644  
Query 872 NGISLGOQTFDGLRLOSQDGRKWAUYTYSAGNDQSPSTVITPDDGFHYHQYQPELD-- 929  
Db 1645 NSI-VTDKTYDANRIKVAERGDVGYT-----YKDFPTQDNK 1682  
Query 930 ---DAVLQVANSB---ITQQFSYNPVTGALLKRAVAGOSLTPYYSGLKKNENIDMK 982  
Db 1683 GKTDKVEVAIINMGDYTAKTSYTYNDLD-----RNRVNDGS 1719  
Query 983 KMSYL-WTLRGLNGYTDLTGTOIKSRDTHGRVTOIKSSIKTTLNVYD-LNRHISQV 1040  
Db 1720 KNAIFEFDEFENINVTYAGNGTAAANYTYDSTQKTYNAIISASQTQLDEYTYDDAASNR 1779  
Query 1041 TDLATGMLTLYVEFDGLNREIGRKLCDSSGHTLDDQSWLKTQOLANRYKLVGLVORT 1100  
Db 1780 TSDINKQDKTTEYDAVNO-----LTKETLPDGYTK----- 1811  
Query 1101 EQSYDSRNRLNQKCDGACPT--DKY--GHSIVTON---FTYDIIGNITACHTPADG 1153  
Db 1812 -AYTYDGFGRKRTQVALISGETKITDASYNQGNOLVSNMGALTYDANGNRT-----SDG 1864  
Query 1154 TEDHATFKANPDPQCLTEVHNTHPMDIRILKYDKAGAVINITDNHGTENFTYDTL 1213  
Db 1865 KYTYT-----WTOGRLSSIT-KKGSSEPT----- 1869  
Query 1214 GRLQNGQSVYVDPLNRVLSQKTDLLDBELVY---RETMVNEVRNGEMIR-LLRATGE 1268  
Db 1890 -----SYTYDDNRRLSKTYDVGTVTHYHGDSDIDLXYETDQKVVYRQVYVSDN 1940  
Query 1269 TTIAGQASAKVLLTGTSQOSVILTSDKQ--LSQEAYSAYGK-----HKSTANDAS 1318  
Db 1941 VRLAMKNNGKTLVYHNAHGDVIALTDBAGKIVAEYVADAGNVLKNTASTEEKKANYG 2000  
Query 1319 ILGYNGBA-----DPVSGVTHLGNGYSYDPTLKRFTPTDLSFPAGAGINPY 1367  
Db 2001 YAGTYTDEIHEQYLLMARIYEPBOV-----FTAYID-----YGBDDP--QTMNGY 2046  
Query 1368 SYCLGDPINSDPSGHLSSW---QAMTIGIMGIAGLLTITATGMAIAAAGSIATAASTS 1424  
Db 2047 NYANNPVPMMFDPDGNVAAWMTAAAGYGAFBEGAEVLLITTKKQNMK-----GFGKAVVUGA 2101  
Query 1425 TTALAFGL 1433

Db 2102 VLGI6GFKL 2110

## RESULT 6

T37218 hypothetical protein SC2H4.02 - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C/Accession: T37218

R/Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Raftery, M.A.

submitted to the EMBL Data Library, September 1998

A/Reference number: 221615

A/Accession: T37218

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Features: 1.2183 &lt;OLI&gt;

A/Cross-references: UNIPROT:O86585; UNIPARC:UPI00000DADB3; EMBL:AL031514; PIDN:CAA20596.

A/Experimental source: strain A3(2)

C/Genetic:

A/Query Match 4.3%; Score 381; DB 2; Length 2183;  
Best Local Similarity 19.8%; Pred. No. 2.5e-12;  
Matches 366; Conservative 221; Mismatches 669; Indels 594; Gaps 91;

```
Qy 99 TGENYKVIETDKYKLOQKK-----LDNLRFKDKLKENCYRIHKSGLIEVLTF 148
Db 620 TGS-----IDQSLMDEIKHTGKRGTDLSLDAVKFSHVMPN--RVDGPSDDI----- 666
Qy 149 NNNAFDLKVPKLLNPAGHAIYIDW-----NFEATQRLNRIYDDLDGHDPL 196
Db 667 --LSFERPRIRTVSSEGAQITVDYLRADCAVQATMPKADENKRCFVYWSPGQSEPI 724
Qy 197 LNEIYQGLITLILFP--GQKEGYTELRLNQLNSINHSIG-----NENPL----- 244
Db 725 LDMFOKRPVSVRTTDLGSEA-----VQHTVEYSGGAGAHVYDDPLTPAKER 773
Qy 245 TMSFGYTRIPKNGILQW-----ITSMTAPG--LKEVTNYSNNGHHFPGSALPVL 296
Db 774 TMS-----OMRGYKVTHTYTPSGGTRATTVTYLRGMGDRVLDGDKTPD 820
Qy 297 PYYTLMKQVPGAGQPAI-----QAEYSYSHNYVGGSGNGIWNKLDMLYGLMTE 346
Db 821 PDARRKAEVSGVTAATDSDQLAGPQRESVAHYGDGEGVNDPMSKST----- 871
Qy 347 YNTGSTRSRYKQEGHDQVRIERTYNNYHLTSCKQONGY--IGTTETAY--YAIIG 402
Db 872 ----ATQHKSYADTEAY--YVRVGAS--HARTITSRLNPDYRVRTTRGYDDYGMV 921
Qy 403 HNPDSQSQFOLPRTKETWRSADNS-----YRSEITETPD--ESGNPL 445
Db 922 SVEDLDGDSVTGDKCTRTWYARNDAGLALVSRTRI VGRACSVTSDALDLPADARP- 980
Qy 446 TKYIKDKTKOKIISPSTHME--YYPAGEVNDCEPEYGFTRFKKIIQTPYSEFPDOP 503
Db 981 GDVVSDDAT---AYDTTMSATQKPTKGA--RMSGRAKYGGDRPLMOTTAATYDY--DTL 1035
Qy 504 EKPIQVRYSLIGSOSHTLKIIEERHYSATOLNLSLFQYNTKSELGRLLKQTECTGEN 563
Db 1036 GRPQVQV---NTDDVITSKTEYQPPAAGPLTSTTV-----N 1069
Qy 564 GKTYSVNKKFTYTKODDTLQOSHITTHDNFTIHRSOVRSRYGRLEFSDPTDKDIYQMS 623
Db 1070 AK-----GHRTTVKDFAL-----GADLKVDANGKVTESA 1100
Qy 624 YDKLGRLLTRTL--NSGTPYANTLLYDYELNNLDONDPFVITTTDVGNGQLRNFPGAG 682
Db 1101 YDLSGRATSVWLPKRSALGKIANYYGYG--VKSSTLPMVSSAT--LNG-----DGG 1150
Qy 683 RHVSQCLKSDGDKGKFTTINTQYDEGGRHNTSTYSYLTNGR----- 725
Db 1151 YRTTYEYDS-----LRTROYAPASQAGRVIAQTLGYDRGLPTAQAADIWDDTAAP 1203
```

```
Qy 726 ----QOTDPKVALSNKSYDNMGOIAN--THMSYGVSE----- 758
Db 1204 AGKIVQIDGGQARQRDSDVYDGMGRVTKAVTYSYGTENMAVDTAARGDLVTGAPBGSA 1263
Qy 759 -KITVDPIITLATAKQLOSNNSNVQSGEVTTTTPSQOPIQITLPDAGHLQSCHTLTRDG 817
Db 1264 NAVVTAFGRTVRRBYAGATQAPAGTYMTTRYA-----FDADROKSTIADNRSA 1313
Qy 818 W-----DRVKEKDAIGQCTIYQYDNRRNIQITLPD 849
Db 1314 WYTYTDLFGQVSVTPDQKTYTEYDALDRAVKSIDGRSEVLLFEYDVLGRKTKMQSA 1373
Qy 850 GTVNRKRYAPSTDTLI-----TDIRNGISLQO-----QTPDGLSRLTQSO----- 891
Db 1374 KTGAN--KLAMSFDTLAKQODPAVAYEGSEGRATQKTRRDPLKYKTNNLPLPAND 1432
Qy 892 ---DGRVWAITYSAG-----NDQCPSTVTTPDQFIHYQYQPELDAVALOVASNEIT 941
Db 1433 PLVAAGVPARLAFSTGYNLDGTVKQAAPAVAGLSAETVSYYT-----DGLQGVLTAKGT 1487
Qy 942 ----QPFSTNPVYGALLKAVAEQSLTPITYPSGRKMEINIDMKKSYLMTLRLGENY 997
Db 1488 TGYLQQAAYS-----LGLRQMTLATDPGAKKVIYLNND--YAGTRRLTRSY 1534
Qy 998 T--DLGTI---OKISRDTHGRVTOIKDSIKTTLNVYDLNRIHISQVDTLATGHMLTTT 1052
Db 1535 VYDDVHGFMLQELAKYQODDAGNITSVDA---TTLG-----GTGKADHOC 1576
Qy 1053 VEEFDGLNREIGRL-----CDSSGHTLDIQQSWMLTQOLANRIYKLVGLQRTQOYSYD 1106
Db 1577 FTYDG--HRLSEAMTPETADCSISGRT-----VAGIGGAAPWTSYQYD 1619
Qy 1107 S---RRRLNOKYDKGAECPD--KYG-----HSI---VT-----QNTFYDIYGNITACHTT 1149
Db 1620 DSGLRKQTEHRNSGDVYTEYEGTAEGQPHLSATVYGAENASYYTDETN----- 1672
Qy 1150 PADGTEDHATFKPAPLTPCOLTEVHHTHPDMDNIRLKTDKARAVINITDNNGTENFT 1209
Db 1673 -----TE---TRGVAAQTGLDMNAGRLAGVSEPPAAGKRPAT 1707
Qy 1210 YDTLGRLONGQSGVGYDPLNRLVSOQKTDLDC--LYRETMVNEVR--NGEM-----I 1261
Db 1708 -----GTAYYADAGDILLIRPPTTDEYVLIGTTEVHLKVSNGAKALSQA 1756
Qy 1262 RLIRTGETIIAQRASKVLLTGDSQOAVLTGDKNLSQEAISA-----YGR 1309
Db 1757 RYKAGSAVIAVYTS---AGVSGTKLTFPLADHGTSGLAINDTLAPAKWSTPFGA 1812
Qy 1310 HKSTANDA--SLGVNGERADPVSGVTHLNGVRSYDPTLMRPHTPDS--LSPGAGGINP 1366
Db 1813 PRGTAGAMPDDRQFGLKPADAAATGLTOL--GARQYDPDGRFLSYDPLLEPKPTLNG 1870
Qy 1367 YSYCLADPINKSDPSGHLISQWAMTGLGMIAGLLTIAT--GGMAIAAGIAAIAASTS 1424
Db 1871 YAYASNSPVTNSDPG-----TSDGLG--GLLGAIGALIGGVAVIGAVGAAL--TA 1919
Qy 1425 TTLALFALSVTSDITSIVSGLALEDASPKRASSILGVNSMGGAAGLAEAIKGTKLATH 1484
Db 1920 VGSILGGGG-----GGGWGCTAPTS--GW-----TQPLTKQMTPGATY 1956
Qy 1485 LGAPADGENALLKS-----TSES-----SRIMGVTRSLDREIVNER----- 1523
Db 1957 -----NFTKSMDLPRNPQSULEMLASMPDWGIYS--DPRKAANWETSRLSPFG 2005
Qy 1524 ----QGVINDHS--RGYTDNFMWKGBOALIVHGDK--DEFLYHTEGNKNG--KGYTRAT 1574
Db 2006 WLMGGGYPAREHODFRG-----GDAFTSLADDETISGLRSWVGQARKGTKGAPAKEV 2059
Qy 1575 PEOUYDLKDNMIVDLTGQGDKEVHLLSCYKSSGAADRAKAKIINRPV----- 1623
Db 2060 GFQYVD-----EGPEPSSPWYKNSLGAKAQIDAGVLTNGAVGTGQADAFLG 2107
Qy 1624 AYSNKPTISQGLARIBRKDPFLKSTYHSYDPRKILIGRTEKTVKPTFRP 1673
```



Matches 223; Conservative 121; Mismatches 308; Indels 327; Gaps 48;

```

Qy 542 YNTDKSLGR---LKTGCTKNGNKTYSVNKFYTTQDDTLQOQSHSTTHDNFTIH 597
Db 32 HHTDSGHOYRLDMNLALERSLCTVDSMGT-----SCHMMAOGLVAY 74
Qy 598 RSQVRSRYTGR-----LFSDDTKDVIQMSYDKLGRLLTTLTNGSTPYANTLTYYDE 650
Db 75 RDEAGCTTTRMSDEBRLLCMTAOGKMYVYDRLGHL-----TETHD-P 120
Qy 651 LNNIQQDNRPPF---VITTDVNGNQLRNEFD-----GAGHVSQCL 689
Db 121 LGRVEQTQMPVHMHPETEVDAAVAVAYEDERGNLAQVSDPLQRTVYGYDNR--CQVV 179
Qy 690 K--DSDDGKRYTHTHQYDEG---RH-----HTSYSDYLNGRQDPPDKVHLSMS 738
Db 180 RITDARGDKYL-----QWNEGQLMHRTDCSGQTMFDERLRLREYDAES--NSTR 232
Qy 739 KSYDNMGQIANTHMSYSEKITVDPITLTATKOLQSNVNVQTEKVTYTPSQOPIQI 798
Db 233 YSYDNGHLEVMADRTERYQD-----AGRLVKTSPAG---QI 272
Qy 799 TLFEDAGHLOCHTLTRDGMVREKETAIGQCTTYQYDNYRVYQITLPDGTIVNRKYA 858
Db 273 TRMQ-----RDGQGRVRRQDGTATGRATAYEDAGRLTTLTNGESYRRRY- 319
Qy 859 PRSTDTLITDRVNGISLGQTPGSLRLTQSP--GGRVAYTYTSAGNDQCP----- 909
Db 320 -----VLDKRVTEQTPGSGRRRYGNALNAVAVYGGERG 356
Qy 910 -----STVTPDQGFHYQYQPLDDAVLY-----ASNEITQOF 944
Db 357 GEIRHGLEBDAAGRLTAKITPE---TREYRYDADRLERRRRHDAEGEEVLR-RF 412
Qy 945 SYNEVTGALLKAVAE-----QSITPIYPSGRLKMNINDMKMSYLTWLTGLEN 995
Db 413 SYDSAGNLSSETAQGLQHRVYDQGNRTETQMPDGR-----TLRYLYX 456
Qy 996 GYTDLTGTICKISDTHGRVYQIIDSSTKITLANDDLNRHIGQVTDLATGMLTTFYEF 1055
Db 457 G---SGHLQOI-----NIGRDVISEFT- 476
Qy 1056 DGLREIGRKLDCSSGHTLDIQGMLKTOQLANRIV--KLNGVLAQT---EQYSDSRNR 1110
Db 477 DMLREYQV---SQGR-LDMRMYDRTGRRLTKLTCGMAGVPEFTIDREYVSGQDE 531
Qy 1111 LMOYKCDGAECPDTKYGHSIVTONFTYDIYGNITAC--HTTPADGTEHDATKFPANPTDP 1169
Db 532 LLK-----KRHSQGVTDYF--YDTGRTITACRNEAVLDSMOYDA--ANLDDR 577
Qy 1170 Q-----LTENHHTHPMPDNIRLKYDKAGVINITDNHGTENFTYDTL 1213
Db 578 QGETAQAAGSVPEFNRLTSYRGLH-----YRYEYGRVVEKRGNG--TOHYRMDAE 628
Qy 1214 GRL-----QNGQGSVYG--YDPLNRLYSOKTDLDCELYRETMLVNEVNGEMIRLLR 1265
Db 629 HRLLEVAVTRGGYTRRGVYTDAPGRVKEKHELDABEGPYRTTFLMDGMLABEGRGR 688
Qy 1266 TGETTIAQOR-----ASKVLTLGTD--SQQSVILTSBKONLSQEA--YSAY 1307
Db 689 SSSLYIYSDRGSHEPLARVDBAARPEADEVLYYHTDVGAPEEMTDGCGNIVMEAGYQVW 748
Qy 1308 GK---HSTANDASILGNGERAPVSGVTHLGNGYSYDPTLKRFRHTPDSLSPFG--AGG 1363
Db 749 GNLTHEKTRVQGNLRFQGYLDRETLAY--MLYRFYDIDIGKIFISGD---PIGLAGG 803
Qy 1364 INPYSYCLDPIINRSDPEG 1382
Db 804 INLYQYA-PNPLSYIDPLG 821

```

RESULT 9  
C64805

thcC protein precursor [similarity] - Escherichia coli (strain K-12)

C:Species: Escherichia coli.

C:Date: 12-Sep-1997 #sequence, revision 17-Sep-1997 #text\_change 09-Jul-2004

C:Accession: C64805; 163400; T88912

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: C64805

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Accession: 1-1397 <BLAT>

A:Cross-references: UNIPROT:P16918; UNIPARC:UPI00001338AF; GB:A5000173; GB:U00096; NID:9

R:Blattner, G.; Gray, J.A.; Kirchmann, J.A.; Lehner, A.F.; Sedocky, A.B.; Vlazny, D.A.;

U. Bacteriol. 172, 446-456, 1990

A:Title: Structure of the rhaA locus from Escherichia coli K-12 and comparison of rhaA w

A:Reference number: 154935; MUID:90094253; PMID:2403547

A:Accession: 169400

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1221-1397 <RES>

A:Cross-references: UNIPARC:UPI000017888F; GB:M29718

A:Note: the sequence is revised in GenBank entry ECOLSCA, release 114, (PIDN:AAC63073.1

R:Oshima, T.; Alba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Honjo, A.; Ikemoto, K.; Inada

T.; Mizobuchi, K.; Mori, H.; Motomura, K.; Nakamura, Y.; Nishimoto, H.; Nishio, Y.; Sai

DNA Res. 3, 137-155, 1996

A:Title: A 718-kb DNA Sequence of Escherichia coli K-12 Genome Corresponding to the 12.7

A:Reference number: 225006; MUID:97061202; PMID:8905232

A:Accession: T48912

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1397 <OSH>

A:Cross-references: UNIPARC:UPI00001338AF; EMBL:D90709; NID:G1651305; PIDN:BA33539.1; P

A:Experimental source: strain K12; Kohara clone 174

C:Genetics:

A:Gene: rhaC

A:Map position: 15.7-16.0

C:Superfamily: rhaC protein

Query Match 3.9%; Score 342; DB 2; Length 1397;  
Best Local Similarity 21.4%; Pred. No. 1,7e-10;  
Matches 256; Conservative 157; Mismatches 382; Indels 404; Gaps 63;

```

Qy 418 KTEFW-RSADNSYRSGETTETTPDESQNPILTKV-----KDKTKQKILS-----PS----- 461
Db 218 RTQTFHREAAGEFGEITGVY--DGAGHFRVLVLTQAQRAEBAQQAISGTEPSAPEDT 276
Qy 462 -----THMEYYPAGEVQNCPEP---YGF-----RFVK 488
Db 277 LPGTYEGRDNGIRLSAVMLTHDEYF-----EULPAALVRYGMYTRGELAAVYDSNT 331
Qy 489 KLIQTPYDSEK-----DDEKEFIQYRYSLIGSOHVTLLKIEBRHSATQLNLSTL 539
Db 332 QVRSFTYDDKRGWVAHRTGRPE--ICRYD---SDGRVTEGLNAGLSY----- 379
Qy 540 FQINTKSELGRLLKQTEC--TKSENGKTYSVNKFYTTQDDTLQOQSHSTTHDNFTIH 597
Db 380 YQEKDRITITDSLRREEVLTQCEGG-LKRVVK-----EHADSVT 421
Qy 598 RSQVRSRYTGRLFSDYTKOIVTQMSYDKLGRLLTTLTNGSTPYANTLTGYELNQLDD 657
Db 422 QSQPDA--VGRKRAQTAAGRTTYSFDDVYGLITRT--TPGGRSAFY----- 468
Qy 658 NRPEFVITTDVNGNQLRNEFDGAGHVSQCLKQSD-----GDKFYTHHQYDEQGRH 712
Db 469 NHHSQLSATGPDLRIIRREYDEWGRLLIGTAPGDITRRYDNPBHDLPQATDAGSR 528
Qy 713 HTSYSDYLNGRQDPPDKVHLSMSKSYDNMGQIANTHMSYSEKITVDPIT-TLRATK 771
Db 529 KMTWMSRY-----GQLSFTDCSGVYTRVYDHRFGQVTAHREBGLSOYRAYDSRQGLIAVK 585

```



Qy 772 QLAGSNMNVQTKKEVITTPSQOPIQITLFDKAGHLOSCHTLTDEGMDRVKETAIGQ 830  
Dy 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG--SRNGTQYDAMGKA 623  
Qy 831 -CTI-----YQTDNKNRVQITLPDGTIVNRKKAPESTDTLTTRVNGISLGQOTF- 881  
Dy 624 ICTTQGGITRSEMDAAGRIVIRLTSNGS-----HTFRYDVLRLIOETGPFDRRTQRYH 678  
Qy 882 -DGLSRLTQSDGGRVMAVYTSAGNDQPSVTITPDQFHYQYQPELDADVLQVANSSEI 940  
Dy 679 HDLTGKLIRSEDEGLV-----THMYD--EADRLTHRTVANGT 714  
Qy 941 TQOFSYNPVTGAL--LKAVAEGSLTPY-YP-GRLEKMEI-----NDM--KXMSY 986  
Dy 715 AERMOYDE-RGMLMDISHISEGHRVTVHYGYDSKGLASEHLTVHNPQTNELMQHETRH 773  
Qy 987 LMTIRGLENGTDLTGITQKISRDTHGNVTOIKSSITKITLNTDNLNHIQSQVTLATG 1046  
Dy 774 AYNAOGLAN--RCIPDSLPVEMWLYG-----SGMLSGMKLG 808  
Qy 1047 HMLTTTVEF--DGLNREIGRKLCDSSGHTLDIQOSWLTQOLANRIVKLANGVLQTEQYS 1104  
Dy 809 D--TPLYETTRDLRHRETLR-----SFRGYELTAYTPAGOLQSO--HLSNLS-DRDYT 858  
Qy 1105 YDSNRRLNQYKCDGACPTDKYHSIVTQNFYDIYGNITACHTTFAD-----GTEDHA 1158  
Dy 859 WNDNGELIR-----ISSPRQ-----TRSYSTGTGRLTGVTHTAANLDIRIPTTDP 906  
Qy 1159 TPKRANPTDPCQLEVNHTTHD-----MPDN-----IRLKYDKAGVINITD----- 1200  
Dy 907 GNRLEDP-----ELHPDSALSMWPNRIADAHVILYRDRHGTLTEKTLPIREGV 956  
Qy 1201 ---NHGNTENFTYDTLGRLOQNGQSVG-----YDPLNRLVSOKTDLDECEL----- 1244  
Dy 957 IRTDDETRHRIHYDSQRLVHTTRQYABPLVESRYLVDLGRYAKVWREREDLTCMM 1016  
Qy 1245 ---YY---RETMVLNE-----VRNGEMIRLR---TSETIIAQA-- 1276  
Dy 1017 SLSRKPQVTWYGMDDBRLTTIONDBRTLIQITQPSFPLLRVETATGELAKTORSLAD 1076  
Qy 1277 ---SKULLGTDSQSVIITSDKQNSQEA----- 1303  
Dy 1077 TLQSGGEDGGSVFPPLVQMLDLRLESEILA--DRVSEBSRKLWASCGLTVAQMSQM 1133  
Qy 1304 -----YSAYGKHSSTANDAS-----ILGNG 1324  
Dy 1134 DPVYTPARKILHYCHDHGRLPLALISTGTTAWIAYEDWENGLNENPHOLOQILIRPG 1193  
Qy 1325 ERADPVSGVTHLGNQYSYDPTLMRFHTPDLSPFG--AGINPYSCIGDPIINSDPSG 1382  
Dy 1194 QQYDEBSGLYY--NMRHYDDPLQGRYITQD--PIGLKGMNFPQYPL-NFISNIDPLG 1246

RESULT 10  
E90658  
RhgG core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain C)Species: Escherichia coli  
CjDate: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
CjAccession: E90658  
RjHayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc A;Reference number: A9629; MUID:21156231; PMID:11258796  
A;Accession: E90658  
A;Status: Preliminary  
A;Molecule type: DNA  
A;Residues: 1-1404 <HAY>  
A;Cross-references: UNIPROT:O8XED9; UNIPARC:UPI00001653A9; GB:BA000007; PIDN:BAB33660.1; C;Genetic(s)  
A;Gene: EC80237  
C;Superfamily: rhaP protein

Query Match 3.8%; Score 341.5; DB 2; Length 1404;  
Best Local Similarity 20.2%; Pred. No. 1.8e-10;  
Matches 255; Conservative 130; Mismatches 370; Indels 505; Gaps 59;  
Qy 376 YHL-LTSECKQONGYIQTETAYAIIGNFDSQPSQFQPKT-KTETMSADNSTYSEI 433  
Dy 245 FHLVLTQAOGRAVFRQKATSSSPGPR--SASSLVEPDLTACTEYGANNGIRLEA 302  
Qy 434 TETTPDES-----GNPLTVINDKTKQKIIISSTHMEYVPAGEV-----DNCPPEPYGF 483  
Dy 303 VMLTHDPAYDELPAADLARYT-----YTASGELAAVYDRSGTQVRGF 345  
Qy 484 TRPVKKIITQPYDSE-----FKDDPEKPIQYRSLISGOSHTVLKIEBRYATOL 534  
Dy 346 A-----YAHNAGRWVAHHYAGRPES--RYRVDGIG-----RVTEL 379  
Qy 535 LN---STLFQYNTDKSELRLKQTEC--TKBENKTYSVNHFYTKKODDTLQOSHST 588  
Dy 380 VNEGGLDYREYQDQRTVITDSINRREVLTEBEGG-LKRVVVK----- 422  
Qy 589 TTHDNFTIHSQVRSRRTGRLFSDDTKDILVTQMSYDKLRLTLRLTNSGTPYANTLTVD 648  
Dy 423 -EHADGSTRSE-----YDEAGRLKAQT----- 444  
Qy 649 YEALNLIQDDNRPPFVITTTDVNGNQLNNEFDGAGRHSQCLKDSDG-----DGKPYT 700  
Dy 445 -----DAAGRREYSLHMAAGAVTAVTGPDGR--T 472  
Qy 701 IHTQOYDEGRHHTSYSDYLTNGROQTPDKVHLSKSYDMMGQIANTHWSYGUSEKI 760  
Dy 473 VR--YGVNSQROVTSVTPDGLARSSREYDEKGRILAETSRSGE-----TRYSYD 520  
Qy 761 TVDPITLTKAQLAGSNMNVQTKKEVITTPSQOPIQITLFDKAGHLOSCHTLTDEGMDR 820  
Dy 521 -DE-----ASELPTGIQDATSGTKQW-----AMSR 544  
Qy 821 ---VRKETDAIGQCTIYQYDNVNRVQITLPDGTIVNRKKAPESTDTLTTRVNGISLG 877  
Dy 545 YGQLTFTDQSGYTTREYDRYGOQLAVHNEBGSITVSYNP-----RG 588  
Qy 878 QQTFDGLSRLTQSDGGRVMAVYTSAGNDQPSVTITPDQFHYQYQPELDADVLQ 934  
Dy 589 Q-----LVSOQDAQGRERTRYESAAGDL--TAIVAPDSRSSEIQY-----DAMGRAVS 634  
Qy 935 VASNETIQPSYNPVTGALKKAVAGOSLTPYIPPSRLKMEININDKMSYLTLAGLE 994  
Dy 635 TTQGGITRSMGYDAAGRITVLTNENGSQSTFRYDPVRLTEQRGFDRGTQRY----- 686  
Qy 995 NGYTDLTGTI-----QKISRDTGHRVTQIKDS-- 1022  
Dy 687 -HYDILGKULQSEDEBEGLVLMHYDASDRITHTRVNGDPRAEQWYDGHGMLTTLSHNSEG 744  
Qy 1023 -IKTTAYDNLNHIQSQ--VTDLATGHWL----- 1049  
Dy 745 HRVSVHYGYDDKGRLTGERQTVENPETGEMLMHEHTGAYSEOGLATROEBDGLPYEWL 804  
Qy 1050 -----TTTVEF--DGLNREIGRKLCDSSGHTLDIQOSWLTQOLANRIVKL 1093  
Dy 805 TYSGGYLAGMKLGSTPLVEYWRDLNREYANSP--GGAVELATWANTSGQLSRHLNL 861  
Qy 1094 NGVLAQTEQYSYDSRNLIQYKCDGACPTDKYHSIVTQNFYDIYGNITACHTTFAD 1153  
Dy 862 P---QDRODDWMDNNGD--IRISPG-----ESREYRSVDQGRLLGVHTTANL 906  
Qy 1154 TED--HATFKRANP--TDPQCLEVNHTTHD-----MPDNIRL-----KYDKAGRYIN 1197  
Dy 907 DIDIPYATDPAQRNLPDP-----ELHPDSTLTAMPDN--RIADAHVYVRYDEYGRILAE 958  
Qy 1198 ITD-----NHGNTENFTYDTLGRL-----QNGQGSV--YGVDP-- 1229  
Dy 959 KTDRIPEGVIRMDERTHAYHSQRLVHTTRIQHGEPQVESRYLVDLGRRTGKRVWR 1018



Qy 1230 -----NRLVSKDTLDELCELYR-----ETMLVNEVRNGEMI 1261  
 Db 1019 RENDLTGMSLSRKREBTWGMDSRLTYQTQOTRQTYVQPSFPLRLRIENKEQA 1078  
 Qy 1262 RLAR-----TGETTIAQ-----QRASKVLLTSTDSQOS-----VILTSKONLS 1300  
 Db 1079 KAHRSRLAEVLQEDTGLTLPALVLMGLRLERLRQGSVEESQOMLQCGLTAEQGAQ 1138  
 Qy 1301 QEA-----YSAVGRKSTANADASI-----LQYN 1323  
 Db 1139 LERGVIPERKLALYHCDORGLPLGLISPRERETALTAYEDENGLSTSAQPLQOSLRFP 1198  
 Qy 1324 GERADPVGYTHLNGYRSYDPTLMRFTPTDLSLSPG-AGGINSYCLADPINSRDPSC 1382  
 Db 1199 GQGVDESSGLVYNNRY--YDPLQGRYITOD---PIGLBGMVLYOYPL-NPIEHIDPLG 1252

## RESULT 11

H91236  
 Rhsp core protein with extension (imported) - Escherichia coli (strain O157:H7, substra  
 C/Species: Escherichia coli  
 C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
 C/Accession: H91236  
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Rep. 8, 11-22, 2001  
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
 A/Reference number: A93629; M01D:21156231; PMID:11258796  
 A/Accession: H91236  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1394 <HAY>  
 A/Cross-references: UNIPROT:Q8X385; UNIPARC:UPI00000029D9; GB:BA000007; PTDN:BA038287.1;  
 A/Experimental source: strain O157:H7, substrain RMD 0509952  
 C/Genetic:  
 A/Gene: ECG4864  
 C/Superfamily: rhsp protein

Query Match 3.8%; Score 341; DB 2; Length 1394;  
 Best Local Similarity 21.5%; Pred. No. 1.9e-10;  
 Matches 260; Conservative 156; Mismatches 404; Indels 392; Gaps 63;  
 Qy 418 KTEETW-ESADNSYRSSEITETTFDESQNPITFKVI-----KDKTKQKIIIS-----PS-----461  
 Db 218 RTQTFHREAAAGFSGSEITGYT-DGAGRHFLVLTQAKRABEQAKISGTEBSAPDT 276  
 Qy 462 -----THMEYPPAGEVNDNCPREP--YGFT-----RFLVK 488  
 Db 277 LPGYTEYGRDNGTRLSAVMLTHDEYP-----ENLPAAPLVRYGWTREGELAVVYDSGK 331  
 Qy 489 KIITQTPDSEFK-----DDEKFIQYRYSLSIGSQSHVTLKIERHYSATQOLNSTL 539  
 Db 332 QVRSFTYDDKYRGMVAHRTGRPE--IRRYD--SDGRVTEQLNPAQLSYT-----379  
 Qy 540 FOYNTDSEKGRLLKQTEC--TKGENKTSYVHKFTYQDDTLQOQSHSITTHDNFTIH 597  
 Db 380 YQIEKDHITTTDSLNRREVLHTQSEGS-LKRVVAK-----EADSGSVT 421  
 Qy 598 RQVRSRYRGLRPSDITDKDITVQMSYDKGLRLTRTANGSTPYANTLTVDYELANLQDD 657  
 Db 422 QSGFDA--VGRLAQDQDAARTEYSPDVVTGLITRT--TPDGRASAVY-----468  
 Qy 658 NRPPVITTTDVANGNQLRNEFDGAGRVASQCLKQSD-----GDKFYTHITQYDEQGRH 712  
 Db 469 NNSQLTSATGPRGLEIRREYDEGRGLIOETAPGDITTRRYNPHSDELCAEDATGSR 528  
 Qy 713 HTSTYSDYLTNGRQDTPDKVHLSSMSKSYNMGQIANTHMSYVSEKTIYDPI-TLWATK 771  
 Db 529 KWTWTSY--GQLSLTSDSGYVTRKYDHRFGMGATVHBEBSQRAYVDSRQGLIAVK 585  
 Qy 772 QLOSNNSNNVQTKGEVTTVTPSQOPIQITLFPDEAGHLQSCHTLTFRDGDWRVKEETDAIGQ- 830  
 Db 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG-SRNGTGYDAMGKA 623

Qy 831 -CTI-----YOYDNNRYVIOITLPDGTIVNRKVAPESTDTLITDRVNGISGQTF- 881  
 Db 624 ICTTQGGTSMREDAAGRVIRLSENGs-----HTFRYVDLDRLLQETGFDGRTORH 678  
 Qy 882 -DGLSRLTQSDGGRVWATYTSAGNDQCPSTVITPPDQFTHYOQEPDLDAVLQVASEI 940  
 Db 679 HDLTKLIREDSGLV-----THMHYD-EADLTTRTVGET 714  
 Qy 941 TQGFSTNPVTGAL-LKAAVSGSLPIYV--PSGRL-KMENINDMKMSYIM-----988  
 Db 715 AERWQYDE-RGMVLTDSHISEGHRVT-VHGYDEKGLTGERQVTHHPQTEALIMQHETR 772  
 Qy 989 ---TLGELNGYTDLTQIKISRDPHGRVTLQKDSIKTTLVNDLNRHIGQVDTLAT 1045  
 Db 773 HAYNAQGLAN-RCTPDSLPVAVEMLYTG-----SGMLAGMKL 807  
 Qy 1046 GNMVTTTVEF--DGLNREIGRKLDSGHTLDIQOSWTKTQOLANRIVKNGVLORTEQY 1103  
 Db 808 GD--TLPVDFTRRLRKTIRRRGRVELTATYPAQLOSGHLNS-----LQYRDY 857  
 Qy 1104 SYDSRNRLNOYKCDGACPTDKYGHSLVTQNTFYDIYGNITACHTTPADGTEDHAFKA 1163  
 Db 858 TAMDNGELIR-----ISSPRQ-----TSYSGSDSGRLTGVTTPAA-----NLDIRIP 900  
 Qy 1164 NPDPDC--QTEVHHHTPD-----MPDN-----IRKTKDAGRIVNITD 1200  
 Db 901 YATDPAGNRLPD-BELHPDSTLSMPDNRJARDAHYLYRDRGRRLTEKTLIPGVIRT 959  
 Qy 1201 NHGNTENFTYDTLGRLONGQ-----GSVGYDPLNLTVEQKTDLLDCEV-----1244  
 Db 960 DDERTRRYHNSQHRLLVHYRTQYEBELVBSRLYDPLGRVAKVARRRDLTGMSLS 1019  
 Qy 1245 -----YV-----RETMVNE-----VANGEMIRLR--TGETTIAQRA-----1276  
 Db 1020 RKPQVTWYGMWDGRLTITQNDRTRIQTIYQPGSFTPLIRVETATGELAKTQSRSLDALQ 1079  
 Qy 1277 -----SKULLTGDSQOSVILTS-----1284  
 Db 1080 QSGGDEGGSVPPFVLVQMLDRLESETLADRVSEBSRWLASGCLTVYQMOQOMDPVYTP 1139  
 Qy 1295 -----DKQNL-----SQEA-----YSAVGRKSTANADASI-----1330  
 Db 1140 AKIHLYHCHDRGLPLALISKGATEKCAEYDWMGNLMBENHOQLRLRQOQYDEE 1199  
 Qy 1331 SGVTHLNGYRSYDPTLMRFTPTDLSLSPG-AGGINSYCLADPINSRDPSCGHLISQAW 1389  
 Db 1200 SGLYY--NRHRYDDPLQGRYITOD---PIGLKGMVLYOYPL-SPVNGMDPLGLYEFKSK 1253  
 Qy 1390 TGIQMGIGAGLL 1401  
 Db 1254 NIDDIGIFALAM 1265

## RESULT 12

B65145  
 Rhsp protein precursor - Escherichia coli (strain K-12)  
 C/Species: Escherichia coli  
 C/Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004  
 C/Accession: B65145; S47701; B36902; A30092; I54935  
 R/Blaetner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A/Title: The complete genome sequence of Escherichia coli K-12.  
 A/Reference number: A64720; M01D:97426617; PMID:9278503  
 A/Accession: B65145  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-1411 <BLAT>  
 A/Cross-references: UNIPROT:P16917; UNIPARC:UPI00001681C5; GB:AE000424; C  
 R/Plunkett, G.  
 submitted to the EMBL Data Library, March 1994

A:Reference number: S47666  
A:Accession: S47701  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1129 / 'O', 1131-1411 <PLU>  
A:Cross-references: UNIPARC:UPI00001338AE; EMBL:U00039; NID:946582; PID:AA818457.1; PI  
A:Experimental source: strain K-12, substrain MG1655  
R:Zhao, S.; Sandt, C.H.; Feulner, G.; Vlazny, D.A.; Gray, J.A.; Hill, C.W.  
J. Bacteriol. 175, 2799-2808, 1993  
A:Title: Rns elements of Escherichia coli K-12: complex composites of shared and unique  
A:Reference number: A36902; MUID:93259920; PMID:8387990  
A:Accession: B36902  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1397-1411 <ZHA>  
A:Cross-references: UNIPARC:UPI000017888C  
A:Note: sequence extracted from NCBI backbone (NCBIN:132073, NCBIPI:132075)  
R:Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.  
J. Bacteriol. 171, 636-642, 1989  
A:Title: rns gene family of Escherichia coli K-12.  
A:Reference number: A91901; MUID:89123133; PMID:2644231  
A:Accession: A30092  
A:Molecule type: DNA  
A:Residues: 1-100 <SAD>  
A:Cross-references: UNIPARC:UPI000017888D  
R:Feulner, G.; Gray, J.A.; Kirchmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;  
J. Bacteriol. 172, 446-456, 1990  
A:Title: Structure of the rnsA locus from Escherichia coli K-12 and comparison of rnsA w  
A:Reference number: 154935; MUID:90094253; PMID:2403547  
A:Accession: 154935  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1221-1411 <RES>  
A:Cross-references: UNIPARC:UPI000017888E; GB:M29717; NID:9147622; PID:9147623  
A:Comment: the rns core consist of two distinct parts: a large N-terminal core that is c  
A:Gene: rnsB  
A:Map position: 77 min  
C:Superfamily: rnsB protein  
C:Keywords: transmembrane protein  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-1411/Product: rnsB protein #status predicted <MAT>  
F:28-55/Domain: transmembrane #status predicted <TM>

Query Match 3.8%; Score 340.5; DB 2; Length 1411;  
Best Local Similarity 20.4%; Fred. No. 2e-10;  
Matches 287; Conservative 172; Mismatches 459; Indels 490; Gaps 70;

Qy 418 KTEW-READNSYSEITETTESGNDPLTKVI-----KDKTQKITS-----PS-----461  
Db 218 RTQTFHRELAHGFSGEITGVT-DGAMRFRVLVLTQADRAEBAQALISGTEBAPFDT 276  
Qy 462 -----THWEYPPAGEVNDNCPPEP---YGF-----RFVK 488  
Db 277 LRGTEYGRDNGIRLSAVMLTHDPEP-----ENLPAAPLVRYGWTPRGELAVVYDRSGK 331  
Qy 489 KIITQPPYSEK-----DDPEKTYQRYSLIGSQSHYTLKIEBHNGATQOLNSTL 539  
Db 332 QVRSFTYDCKYGRWVAHRHTGRPE--IRYRD---SDGRVTEQOLNPAGLST-----379  
Qy 540 FOYVNDKSELGRLLKQTEC--TKGNGKTVSVHFKFTYTKODDTLQSHSITHDNFTIH 597  
Db 380 YQYEDRTTITDSDRREVLHTQAG--LKRYYK-----EADGSVT 421  
Qy 598 RSQVRSRYTGRLLFSDDTKDIYQMSYDKLGLLTLRTINSCTPYANTLLTYDELANLQDD 657  
Db 422 QSQFDA--VGRRAQTDAAGRTTEYSPDVATGLITRIT---TPDRASAIFYNNHNOQ---473  
Qy 658 NRPPVITTTVDNNGQALNEFPDAGRHVSQCLAKSD-----GDGKFTYTHIQVYQGRH 712  
Db 474 -----LTSATGPDGLREYDELGRLLIQTAPDDITRRYRDNPHSDLPCTATEATSSR 528  
Qy 713 HTSTVSDVLTNROQOTDDPKVHLSMSKSYDNWGOALNTHWSYGVSEKITVDPI--TLTATK 771

Db 529 KMTWRSRY---GQLLSFTDCSGYVTRYDHRFGQMTAVHREBSGLQRAYADSRGOLIAVK 585  
Qy 772 QLOSNENNVOYTGKAVTTPPSQOPITQTLFDEAGHQ-----SGHTLLTRQMDRVRKE 824  
Db 586 DTGSHETRYE-----YNIAGDLTAIVAPDSRRGTGYDAMGKVRT 626  
Qy 825 TDAIGCTTYQYDNVNRVIOITLPDGTIVNRKAPSTDTLTDIRVNGISLQQTPE--D 882  
Db 627 TQG-GLTRSENYAAGRVIRLTSENS-----HTTRVYDLIDLQIGTGDGRQORHND 680  
Qy 883 GLRLTQSDQDGRVMAVYVSAGNDQCPSTVITPDQGFHYQYQPELDAYLQVANSNETQ 942  
Db 681 LTGKLIRSEDEGLV-----THWYD-EADRLTRFTVKGSTAE 716  
Qy 943 QFSTNPTTGL--LKAVBEGSLTPITY---PSGR--KXENINDMKMSYLV-----988  
Db 717 RMQYDE-RGWLTLSHISEGRVA-VHYRYDEKGRLLGERQYVHPQTEALMLQHETRA 774  
Qy 989 -TLRGLNGYTDLTGTQIKISRDTGRVTOIKDSIKTTNLYDNLNHISSQVTDLATGH 1047  
Db 775 YNAQGLAN--RCLPDSIPAVEMLTG-----SGYLAGKGLD-----809  
Qy 1048 MLTTFVEF--DGLNREIGRLCDSSGHTLIDIQSWLTKQOLANRYKLVGLQTEGYSY 1105  
Db 810 --TPLYEYTRDRLHRETLR-----SFGRYELTAYTPAGOLOSQ--HLNLSLS--DRDYTW 859  
Qy 1106 DSRRLNLYQCDGAECTDKGHSIYQNTFYQNTIYQNTACHTFADGTEDHAFKPNP 1165  
Db 860 NDNGELIR-----ISSRQ-----FRSYSTGTGLVHTTAA-----NLDIRLPYA 902  
Qy 1166 TDPC--QUTEVHHHTPD-----MPDN-----IRLKYDAGRIVNITD-----NH 1202  
Db 903 TDPAGNRLPD-PEIHPSITSLMWPDRRIARDANILVYDNGRLTERTDILPEGVIRTD 961  
Qy 1203 GNTENFTYDTLGRLONGQ-----GSYYGIDPLNRLVSQKTDITDCEI-----1244  
Db 962 ERTIRHYDSQHRLVHTRTQYEEPLVESRYLYDPLGRVAKRWRRERDITGMWSLSRK 1021  
Qy 1245 ---Y-----REMLVNE-----VRNEMIRLIR-----TGRITIAQORA-----1276  
Db 1022 PQVTWYMGDGRLLTTONDSRIQTIYQPSFTPLINVEATAGBELAKTORSLADALQOS 1081  
Qy 1277 -----SKVLLTGTDSQGSVILTSDKONLSQEA-----1303  
Db 1082 GGEKGSVVPVPLVQMLDLRESILA---DRVSEBRMLASGLTVEGKQKQMDPYVT 1138  
Qy 1304 -----YSAVGNKSTAND---ASITGVNGERADP 1329  
Db 1139 PARKIHLVHCDHRLPLALISTEGATWCAEYDEWGMILNENPHOLOQLIRLPGQCYDE 1198  
Qy 1330 VSGITHLGNCRSVDPPLMRPHTPDLSLSPRG--AGINPYSCLDGPTNRSPPSGHLSMQA 1388  
Db 1199 EBSGLY--NRHRYVDPLQGRITTD---PIGLKGMVLYGQL--NPISDIDPLGLSMWED 1252  
Qy 1389 WTGIGMGIAGLLTITANGMAIAAGIAAIAASTSTTALAFCALSVTSIDTISVSGALE 1448  
Db 1253 -----AASG--ACTNGLCGLSA-----MGPDFFDSIDSTAY 1283  
Qy 1449 DASPRASIIIGWVSMGGAAGLASAIGKTYKLATHLGAPEDEGNNALIKSTSSSRILKW 1508  
Db 1284 DALNKINS-----OSICEDKEFA-----1301  
Qy 1509 GVTSSLBREIYRNEGOVYIKDHSRGYTDNFMWKGEO-----ALVYGDCKGFLYHTEG 1561  
Db 1302 -----GLICKNSGRYFSTADNRGRKGSYFPNPSCPNGTEKRVASVYHTHG 1346  
Qy 1562 NKHNGK--GPYTRATPEOLVDYLDKNNI 1587  
Db 1347 ADSHGTYWDELFPSGKDEKIVK--SKDNNI 1373

RESULT 13





C/Gene: rhea  
 C/Superfamily: rhae protein  
 C/KeyWords: transmembrane protein  
 F/1-26/Domain: signal sequence #status predicted <SIG>  
 F/23-1377/Product: rhad protein #status predicted <MAT>  
 F/28-55/Domain: transmembrane #status predicted <TM>

Query Match 3.7%; Score 331.5; DB 2; Length 1377;  
 Best Local Similarity 20.8%; Pred. No. 6.2e-10;  
 Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;

211 LPPQKEGY-RTETLPIRLQINSIHNSLGNENPLTWSFGYTPGKNGILGQWITSTAP 269  
 126 LPPGE-DGYSRSESLVLR--GGVAKDEGHRLLALNQ----- 160  
 270 GGLKETVNYG-----NNNG-----HHFPOSANL--PVLPLYTLMKQVPGAGQ 310  
 161 -ALPEELRLSPHRYLATNSPQSPWMLGWCERPEADEVLPAFLPYRVLGLVDRFGRT 219  
 311 PAIQAEYSYSHNVGSGNSGIMNNKDLNGLMTENVGSTESEKRYKDEGHDQIVRIE 370  
 220 QTFHRELAGESGEITGVTDGAMRH-----FRLVL-----TTOQR----- 255  
 371 RTYNNHILTSCKQNGVIQTETAYATIGHNFDSPSOFOLEPKT---KTETWEGADN 427  
 256 -----AEKAO-----AISG---GTBPNAF--PDLPGYTEYGR--DN 287  
 428 SYSEITEITPDESNGPLTKVINDKTKOKIISPSTHWEYPPAGEVDCPEP---YGFT 484  
 288 GIRLSAVMLTHDPE-----YP-----ENLPAALVRYGWT 317  
 485 -----RKYKTIOTPYDSEFK-----DDEPKIORYSLISQSHTLKIE 525  
 318 PRSELAVVYDSGQKVSFTYDDCKRGMVAHRRYGRPE--IRRYD---SDGRVTEQLN 372  
 526 ERHYSATOLNSTLFOYNTDKSELGRLLKQTEC--TGKNGKTVSVVHKFTYKQDDTLQ 583  
 373 PAGLSY-----YQEKDRITITDLSLDRREVLTGEGAG-LKRVYK----- 413  
 584 QSHSITTHDNFTIHRSGVRSGRYTGRFSDTDTKDVIYQMSYDKLGRLLITRLNSGTPYAN 643  
 414 -----EHADGSVYQSQFDA--VGRLAQOTDAAGRTTEYSPDVVTGLITRIT--TPDGR 462  
 644 TLTYDYELANLQDDNRPPIVITTDVANGNLNRPDAGRHVSQCLKQSD-----GDGKF 698  
 463 ASAFYYNNHNO-----LTSATGPDGLRLRREYDELGRLLIOTAPDGDITRYRYDNP 514  
 699 YTIHQOYDEQGRHHTSTYSIDYLTNGROQTPDKVHLSMSKSYNMGOIANTHWSYGVSE 758  
 515 SDLPKATEDATGSKKTTWRSY---GQLSFTDCSGYTRYDHDHDFGOMTAHVHREGLSQ 571  
 759 KITVDP--TLTATKQLOSNNSNNVQTGKEVTTYTPSQOPIQITLPEAGHLQ-----SC 810  
 572 YRAVDSRGOILAVKDTQGHETRYE-----YNAGDILAVIAPDGS 612  
 811 HTLTRGMDKVRKETAIGCCTIYQDNNRVIOITLPDGTIVNRKTAPESTDTLTDIR 870  
 613 NGTOYAMGAVALRTQG--GLTRSMEDYDAAGVIRLITSENGS---HTTFRYDVLDRLLQ 666  
 871 VNGISLQOQTF--DGLSLRTOSODGGRVMAVYASAGNDQCPSTVITPDQFHHYOQPEL 928  
 667 ETGFDGRTQRYHNDLTGKLIRSEDEGLV-----THWHD-BA 702  
 929 DDAVLQVANSNETIQQFSYNPVYTGAL--LKAVALGOSLPIYU---PSGR--KXENINDM 981  
 703 DRLTRTVKGETAERWQYDE--RGWLTDISHISEGRVA--VHYRYDEKGRLTGGRQIVNHP 760  
 982 KKSISYLM-----TLKGLENGITDLTGITQKISDTHGRVTOIKDSIKTTILNYDDL 1033  
 761 QTEELLMQHEHTRHAAYNAQGLAN--RCIPDSLPAVEMLYG-----SGYLAGKCLGD-- 809  
 1034 RHIGSQVTLATGMLTTFVEF--DGLNREIGRKLCDSSGHTLIDQOSWLTQOLANRIV 1091

810 -----TPLVEYTRDRRLHRETL-----SGRYEYLTAATPAQLOSG-- 846  
 1092 KLVGLQRTQYQYDSNRRLNQKCDGACPTDYGHSIVYQNTFYDIGNITACHTTFA 1151  
 847 HNLNLS--DRDYTWMDNGELIR-----ISSPRQ-----TRSYSTGTGLTGVHTTAA 893  
 1152 DGTEDHATPKFANPTDPC--QLTEVHHHTHPD-----MPDN-----IRLKYDAGRVIN 1197  
 894 -----NLDIRIPYATDPAQRLPD--PELHPDSTLSMPDNRIADAHYLYRYDRHGLTE 947  
 1198 ITD-----NHGNTENFTYDGLRLONGQ-----GSVYGYDPLNRLVSOKTDT 1239  
 948 KTDLIPRGVARTDERTHRHYHDSQHRLVHTTQYEEPLVESRYLVDLGRVAKVWR 1007  
 1240 LDCEL-----YI-----RETMVLNE-----VRNGEMIRLIR---TGETI 1270  
 1008 RERDLTGMSLSRKPQVLTWGMDDRLTTIQNDRTRIQTIIYQSGSFPLIRVETATGELA 1067  
 1271 IAOORA-----SKVLLGTDSQGSVL-----TS 1294  
 1068 KTORSLADALQSGGEGGSVVPVLYQMLDLRLSEILADRVSESRMLASGGLTVE 1127  
 1295 DKON-----LSQEA-----YSAYGKHSTAND--AS 1318  
 1128 QMONQMDPVYTPARKIHLHYCHDRGLPLALISKEGTEMCAYDEWGNLNEBNPHQLQ 1187  
 1319 ILGNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFG--AGGINPYSYCLGDPINR 1377  
 1188 LIRLPQQYDEBSGLVY--NRHRYVDPLOGRYITQD--PIGLKGGWNPYQYPL--NBYTN 1241  
 1378 SDPSG 1382  
 1242 TDPG 1246

Search completed: January 30, 2006, 09:49:50  
 Job time : 40.1754 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2006, 09:40:48 ; Search time 133.373 Seconds  
(without alignments)  
8849.971 Million cell updates/sec

Title: US-09-889-874A-23  
Perfect score: 8879  
Sequence: 1 VYIKFLKFRIRITMSDNNF.....PRKIIIGRTKTVKPKTRFP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: UniProt 05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8815	99.3	1660	09EVR7_XENBV	09EVR7 xenobacillus
2	1632	18.4	1625	07MB38_PHOIL	07mb38 photorhabdus
3	1615	18.2	1578	04ZUT5_PSEBY	04zut5 pseudomonas
4	1614	18.2	1562	0883V9_PSEBY	0883v9 pseudomonas
5	1598	18.0	1385	07NW2_CHRVO	07nw2 chromobacte
6	1583	17.8	1775	07NA47_PHOIL	07na47 photorhabdus
7	1553	17.5	773	09S6J1_COXBU	09s6j1 coxiella bu
8	1530.5	17.2	774	052880_COXBU	052880 coxiella bu
9	1490.5	16.8	1682	04ZUT9_PSEBY	04zut9 pseudomonas
10	1416	15.9	1632	088LP7_PSEBY	088lp7 pseudomonas
11	1214	13.7	709	045948_COXBU	045948 coxiella bu
12	1037	11.7	1290	088LP2_PSEBY	088lp2 pseudomonas
13	932.5	10.5	982	0883W6_PSEBY	0883w6 pseudomonas
14	920.5	10.4	1669	087V6_PSEBY	087v6 pseudomonas
15	919.5	10.4	505	045905_COXBU	045905 coxiella bu
16	919.5	10.4	526	09X626_COXBU	09x626 coxiella bu
17	916.5	10.3	526	052883_COXBU	052883 coxiella bu
18	912.5	10.3	528	045949_COXBU	045949 coxiella bu
19	884	10.0	589	045902_COXBU	045902 coxiella bu
20	648	7.3	762	0931I8_VIBIR	0931i8 vibrio pom
21	552	6.2	370	07NVR8_CHRVO	07nvr8 chromobacte
22	486	5.5	389	07NVA0_CHRVO	07nva0 chromobacte
23	478.5	5.4	310	045904_COXBU	045904 coxiella bu
24	450.5	5.1	2217	08TP72_METAC	08tp72 metacoccar
25	450	5.1	380	07N05_CHRVO	07n05 chromobacte
26	439	4.9	2554	0720R5_LEPIC	0720r5 leptospira
27	432	4.9	2334	WAPA_BACSU	077833 bacillus su
28	429.5	4.8	336	045946_COXBU	045946 coxiella bu
29	420.5	4.7	2364	082R58_STRAW	082r58 streptomyce
30	418.5	4.7	1835	06MD86_PARUV	06md86 parachlamyd
31	414.5	4.7	2221	073B23_BACCI	073b23 bacillus ce

32	405.5	4.6	2221	2	Q4MM58_BACCE	Q4mm58 bacillus ce
33	403	4.5	371	2	Q4ZK7_PSEBY	Q4zk7 pseudomonas
34	396	4.5	1513	2	07NY44_CHRVO	07ny44 chromobacte
35	395	4.4	820	2	082OR1_STRAW	082or1 streptomyce
36	395	4.4	2246	2	073C66_BACCI	073c66 bacillus ce
37	393.5	4.4	1976	2	08DAR5_VIBVU	08dar5 vibrio vuln
38	393	4.4	363	2	Q4ZLS4_PSEBY	Q4zls4 pseudomonas
39	392	4.4	1250	2	082RX1_STRAW	082rx1 streptomyce
40	391.5	4.4	2370	2	082RE3_STRAW	082re3 streptomyce
41	391	4.4	391	2	Q4ZK8_PSEBY	Q4zk8 pseudomonas
42	388.5	4.4	1826	2	0987Z7_RHILU	0987z7 rhizobium l
43	386	4.3	2167	2	092BK5_LISIN	092bk5 listeria in
44	386	4.3	2224	2	081U00_BACAN	081u00 bacillus an
45	382.5	4.3	1117	2	08CK70_STRCO	08ck70 streptomyce

## ALIGNMENTS

RESULT 1									
ID	Q9EVR7_XENBV	PRELIMINARY	PRT	1660	AA				
AC	Q9EVR7								
DT	01-MAR-2001 (TREMBlrel. 16, Created)								
DT	01-MAR-2001 (TREMBlrel. 16, Last sequence update)								
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)								
DE	Nematocidal protein 2.								
GN	Name=xnp2;								
OS	Xenobacillus bovienii.								
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;								
OC	Enterobacteriaceae; Xenobacillus.								
OX	NCBI_TaxID=40576;								
RN	[1]								
RP	NUCLEOTIDE SEQUENCE.								
RC	STRAIN=173;								
RA	Morgan J.A.W., Quesley M., Ellis D., Jarrett P.;								
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.								
DR	EMBL: AJ296651; CA19493.1; -; Genomic_DNA.								
DR	InterPro: IPR006530; YD.								
DR	Pfam: PF05593; Rhs_repeat; 5.								
DR	TIGRFAMs: TIGR01643; YD_repeat_2x; 4.								
SQ	SEQUENCE 1660 AA; 185738 MW; AFABA20AD70B164 CRC64;								
Query Match 99.3%; Score 8815; DB 2; Length 1660;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	14	MSDNNFEFTQANNFTSAVSGVDPRGVLNYIQITLGHIVGNLGPPLPLSYSPLNKT	73						
DB	1	MSDNNFEFTQANNFTSAVSGVDPRGVLNYIQITLGHIVGNLGPPLPLSYSPLNKT	60						
QY	74	DIFRGIGFNGFGLSVYDRKNSLSLSTGENYKVIETDKTVLQOKLDNLPKDKLNCY	133						
DB	61	DIFRGIGFNGFGLSVYDRKNSLSLSTGENYKVIETDKTVLQOKLDNLPKDKLNCY	120						
QY	134	RIIHKSQDIEVLGPNNAFADLVKPKLLNPAHAITYIDNFPATOPRLRIYDDLDGHD	193						
DB	121	RIIHKSQDIEVLGPNNAFADLVKPKLLNPAHAITYIDNFPATOPRLRIYDDLDGHD	180						
QY	194	IPLLNLEYQGLIKITLITLPGQKEGYRTELRLNROLNTHNLSLGENDLTWSFGYTP	253						
DB	181	IPLLNLEYQGLIKITLITLPGQKEGYRTELRLNROLNTHNLSLGENDLTWSFGYTP	240						
QY	254	GKNGIIGQWITMTAPAGLKEITVYNNQGHHPPOSANIPVLPIYVTLKQVPGAGPAI	313						
DB	241	GKNGIIGQWITMTAPAGLKEITVYNNQGHHPPOSANIPVLPIYVTLKQVPGAGPAI	300						
QY	314	QAEYSTSHYVGGSGNGIWNNDLNDYGLMTEYVNGSTESRRYKXEGHDOIVRIERTY	373						
DB	301	QAEYSTSHYVGGSGNGIWNNDLNDYGLMTEYVNGSTESRRYKXEGHDOIVRIERTY	360						
QY	374	NNYHLTSECKQONGYIQTTETAYVAILIGHNFDSPQFOLPPTKTETWRSANNSYSEI	433						

Db 361 NNHLLTSECKQNGYIQTETETAYAILGNFDSQPSQFOLPKTETETWBSADNSYRSEI 420  
Qy 434 TETTFDSGNPLTVIYIDKTKOKIISPSTHEVYYPAGEVONCPPEYGFTRPFKXIIOT 493  
Db 421 TETTFDSGNPLTVIYIDKTKOKIISPSTHEVYYPAGEVONCPPEYGFTRPFKXIIOT 480  
Qy 494 PYDEFPKDDPEKFIQYRYSLIGSOSHVLKLEERHVSATQULANSTLFOYNTDKSELGRLL 553  
Db 481 PYDEFPKDDPEKFIQYRYSLIGSOSHVLKLEERHVSATQULANSTLFOYNTDKSELGRLL 540  
Qy 554 KQTECTYGENGKITYSVVHKFTYTKQDITLQOSSHITTHDNFTHRSQVRSYTRLSYDT 613  
Db 541 KQTECTYGENGKITYSVVHKFTYTKQDITLQOSSHITTHDNFTHRSQVRSYTRLSYDT 600  
Qy 614 DTQOIVTQMSYDKIGRLITRTLNSTGPYANLTVYDEYLNLODNRPPEVITTTDVGNO 673  
Db 601 DTQOIVTQMSYDKIGRLITRTLNSTGPYANLTVYDEYLNLODNRPPEVITTTDVGNO 660  
Qy 674 LRNEFDGAGRHSOCLKDSGDGKFTYITHTQOYDEQGRHHTSYSDYLTNGRQOTDPDKV 733  
Db 661 LRNEFDGAGRHSOCLKDSGDGKFTYITHTQOYDEQGRHHTSYSDYLTNGRQOTDPDKV 720  
Qy 734 HLSMSKSYDNMGQIANTHWSYGVSEKITTVDITLTAIKOLOSNNSNNVQTEKVTYTPSQ 793  
Db 721 HLSMSKSYDNMGQIANTHWSYGVSEKITTVDITLTAIKOLOSNNSNNVQTEKVTYTPSQ 780  
Qy 794 QPQITTLFDEAGHOSCHTLTRDMDRVRKETAIGCTIYQYNNVNRVIOITLPDGTIV 853  
Db 781 QPQITTLFDEAGHOSCHTLTRDMDRVRKETAIGCTIYQYNNVNRVIOITLPDGTIV 840  
Qy 854 NRKTAPESTDTLITDIRVNGISLGOQTFDGLSRLOSQDGRWAAITYSAGNDQCPSTVI 913  
Db 841 NRKTAPESTDTLITDIRVNGISLGOQTFDGLSRLOSQDGRWAAITYSAGNDQCPSTVI 900  
Qy 914 TPDCQFIHYQYQPELIDAVLOVANSNETIQOQSYNVPVTKALKAVAEQSLFTPIYPSGR 973  
Db 901 TPDCQFIHYQYQPELIDAVLOVANSNETIQOQSYNVPVTKALKAVAEQSLFTPIYPSGR 960  
Qy 974 KMEINIMKMKSYMTLRGLNGVYTLTGCTIOKISRDTGHRVTOIKOSIKTTLYNDYDLN 1033  
Db 961 KMEINIMKMKSYMTLRGLNGVYTLTGCTIOKISRDTGHRVTOIKOSIKTTLYNDYDLN 1020  
Qy 1034 RHISQOYVTLATGHTLTTVEFPDGLNREIGRKLCDSSGHTLIDIOQSWLTKOQLANRYKL 1093  
Db 1021 RHISQOYVTLATGHTLTTVEFPDGLNREIGRKLCDSSGHTLIDIOQSWLTKOQLANRYKL 1080  
Qy 1094 NGVLOFTEQOYSYDSRNRLNOYKCDGACPTDKYGHSHIVTONFTYDIYGNITACHTTPADG 1153  
Db 1081 NGVLOFTEQOYSYDSRNRLNOYKCDGACPTDKYGHSHIVTONFTYDIYGNITACHTTPADG 1140  
Qy 1154 TEDHATPEFANPTPCQULTEVHHTHPMDNIRLKYDAGAVINTDNHGTENFTYDTL 1213  
Db 1141 TEDHATPEFANPTPCQULTEVHHTHPMDNIRLKYDAGAVINTDNHGTENFTYDTL 1200  
Qy 1214 GRLQNGQSVYGYDPLNRLVSOQTDITLCELYYRETMVNEVRNGEMIRLRTGTTIIAQ 1273  
Db 1201 GRLQNGQSVYGYDPLNRLVSOQTDITLCELYYRETMVNEVRNGEMIRLRTGTTIIAQ 1260  
Qy 1274 QKASVLLTGTDSQSVILTSQKONLSQEAYSAYGKHSSTANDASILLGYNGERADPYSGV 1333  
Db 1261 QKASVLLTGTDSQSVILTSQKONLSQEAYSAYGKHSSTANDASILLGYNGERADPYSGV 1320  
Qy 1334 THLNGVNSYPTLMRFHTPDSLSFPGAGINPYCYCDPRLNDSBPGHLSQWQMTGIG 1393  
Db 1321 THLNGVNSYPTLMRFHTPDSLSFPGAGINPYCYCDPRLNDSBPGHLSQWQMTGIG 1380  
Qy 1394 MGIAGLITLTATGMAIAAGGIAAIAASTSTTALAFALSTYSITISIVGALADAPK 1453  
Db 1381 MGIAGLITLTATGMAIAAGGIAAIAASTSTTALAFALSTYSITISIVGALADAPK 1440  
Qy 1454 ASSIIGVNSMGGAAGLAESAIGKGTKLATHLGAFAEDGENALLKSTSESSRIKMGVTRS 1513  
Db 1441 ASSIIGVNSMGGAAGLAESAIGKGTKLATHLGAFAEDGENALLKSTSESSRIKMGVTRS 1500

Qy 1514 LDREIVANREGOYIKDHSRGYTDNFMKGEOALILVHGDXDGFYHTEGNKNGKGPYTRH 1573  
Db 1501 LDREIVANREGOYIKDHSRGYTDNFMKGEOALILVHGDXDGFYHTEGNKNGKGPYTRH 1560  
Qy 1574 TPQOLVDYLKONNIVDLTQGGDKPVHLLSCYKSSGAAIDMAKINRPVLAISKPTISQ 1633  
Db 1561 TPQOLVDYLKONNIVDLTQGGDKPVHLLSCYKSSGAAIDMAKINRPVLAISKPTISQ 1620  
Qy 1634 GLARIEKDFELKSTYASVDPKRIILGRTKTYKPKTRP 1673  
Db 1621 GLARIEKDFELKSTYASVDPKRIILGRTKTYKPKTRP 1660

RESULT 2  
Q7MB38\_PHOHL  
ID Q7MB38\_PHOHL PRELIMINARY; PRT; 1625 AA.  
AC Q7MB38;  
DT 01-MAR-2004 (TEMBLrel. 26, Created)  
DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)  
DE Similar to the nematocidal protein 2. Probable membrane protein.  
GN OrderedLocustNames=D1u2222;  
OS Photobacterium luminescens (subsp. laumondii).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photobacterium.  
OX NCBI\_Taxid=141679;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=TT01;  
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;  
RA Duchaud E., Ruenick C., Frangeul L., Buchrieser C., Givaudan A.,  
RA Taourit S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,  
RA Daes E., Derose R., Derzelle S., Freyssinet G., Gaudriault S.,  
RA Medigue C., Lanois A., Powell K., Siguler P., Vincent R., Wingate V.,  
RA Zouine M., Glaeser P., Boemare N., Danchin A., Kunst F.,  
RT "The genome sequence of the entomopathogenic bacterium Photobacterium  
RT luminescens.";  
RL Nat. Biotechnol. 21:1307-1313 (2003).  
DR EMBL, BX571866; CAEL4515.1; -; Genomic\_DNA.  
DR Photobacter; p1u2222; -.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RNS\_repeat; 7.  
DR TIGRFAMs; TIGR01643; YD\_repeat\_2x; 3.  
KW Complete proteome.  
SQ SEQUENCE 1625 AA; 180118 MW; BFA362F4BCB3BA97 CRC64;

Query Match 18.4%; Score 1632; DB 2; Length 1625;  
Best Local Similarity 31.0%; Pred. No. 1,48-79;  
Matches 494; Conservative 245; Mismatches 653; Indels 204; Gaps 59;

Qy 18 NEFTQANNFTSANSGVDPRTGILYNQITLGHIVGN--GNLGPFLPLTSLYSPLNKTDI 75  
Db 4 NDIYSNAPNFQSYINTGVDPRTGYSANINITLRPNVNGVIGOV--LNISSFPLTILNN 61  
Qy 76 GFQIGFNGFSVYDRKNSLSLSTGENYK---VLETDKTVLQOKLNDLRFERKOLKENC 132  
Db 62 GFQIGFNGFSVYDRKNSLSLSTGENYK---VLETDKTVLQOKLNDLRFERKOLKENC 120  
Qy 133 YRIHKSQDIEVLTGFNNNAFDLKVPKLLNPAGHAIVIDWNEFATOPRLNRIYDLDGH 192  
Db 121 FYVYNKNGIIEFTLKRIGSS--DIKTVALEFPDQ-----EVPDLIYNS 161  
Qy 193 DPLINLEYQGLITITLTL-FPGQKEGRTETRL--NRQLSINHPFSGNENPLTWSCY 250  
Db 162 RFALSEIKRYMTGTIYKLKNSGNN---CTSVETPDNNISAKIAFYDRNDYLLTVTPY 218  
Qy 251 TPICK-----NGILGOWITSMTAPGLKETVNSNNNOGHFPQSANTLPTLP 297  
Db 219 DASGPIDSAFKKTYQTLKGI-FVISAFTPTGYELVLSYKEN--GKATVDTESIFPA 275  
Qy 298 YVTLMKQVPGAQPAQAEYSYTS--HNYVGGSNGI--MNNKLDNLXGLMTEYNTGSTESR 355  
Db 275 YVTLMKQVPGAQPAQAEYSYTS--HNYVGGSNGI--MNNKLDNLXGLMTEYNTGSTESR 355



```

Db 276 ALTQ---PANGOPAIKSEYESSVHNPLGYSGRTSPDSSQDNLVLTGKATYSIE-- 330
Qy 356 RYDKGEGHDQIVIERITNNYHLLTSECKOONGIQTTERAYVAIGHNDSQSQQLP 415
Db 331 --RYLNOQNYISYERVFDFKHLMTKAKQDNKRRIITETIYNBPKSPSEOPENLOP 388
Qy 416 K-----TKTEIWMRSANYSRSEITETTPDESQNPILKVIKDKTKOKIISPTWMEYYP 469
Db 389 SHVLTRTDIDQT-----NTSRBSVNIKSDWMGNTL-ITETSGIQK-----EYVYFV 436
Qy 470 AGEVNCPPPEPYGTFREPKKIIQTPYDSEPKDPEKEIYRYSLIGSQSHVTLKIEERHY 529
Db 437 NGEANNCPADPLGFSRFLKSVTGKSPDAQSVANRVTSYQGLPFTGAHYK--EYVS 494
Qy 530 SATOLANSTL---FOY---NTDKSELRLKQTECTGKNGKITYSVHAKTTYKODDTLQ 583
Db 495 KASBETIDSKIVREFNYNSPFTNKS--GSLAKITSVMNNO-----QYTTTKEYESDEMT 548
Qy 584 OSHSITTHDNFTIHRSGVRSRYTGRLPSPDPTKDIYOMSYDKGRLLTTLNAGTPYAN 643
Db 549 TNGTVPDGTGHWESKAVTISITTRKQLRKVDVNHVITDQSYDLSGRITIGQIIDGTIKET 608
Qy 644 TLTYDELANNLODNRPPIYITTDVANGQLRNEFDAGRHVSQCLKDSG-----D 695
Db 609 KRSTIYQYPGDENDFWP-VWIEIDSGIRKTHYDGMGRICIEQDDDGWGTSGIYQ 667
Qy 696 GKTYTITTOQYDEGGRHHTSYSDYLTN-----GRQOTDPKXHLNKSISYDNGQIANT 750
Db 668 GYRKRVLARQYDVLGQLVKESISNDMWLDSANPLRLTTP--LVTTITYQYDGMKNYST 725
Qy 751 HMEYGVSEKTIPTPITLTATK-----QLOSNNNVOTGKAVTTPYSQOPIOTLTDEA 804
Db 726 EYSDGRLEIHPITITTIQGVKGLMNIQNNF-----EGPASKIVYPPD 773
Qy 805 GHLQSCHTLRDGMWRKETAIGQCTIYQYDYNVNIQITLPDGTIVNRKAPFSTDT 864
Db 774 GAIYSTRYRYDGRVTEITDAGVATQIEYDFRIVKTLIPDRITLESVAVSFHEE 833
Qy 865 LIIDIRVNGISLGQOTDGLSRLTQSDGGRWATYSAGNDQCPSTVITP--DQGITIH 922
Db 834 LISALNANGQLSLVYDGLGRVTRDVGKRTKTEYLIGSQODK-PISQVTPAHKKONIDY 892
Qy 923 QYQBELDAVQVANSNEITQOFSYNPVTGALLKKAIVAGOSLTP--YYPSGRLLKMNEN-D 980
Db 893 LY--ALQVMSKFTTETSQONFSYQKTGALLSA-TGVQSOSNYSYPSGYLQIESSRD 949
Qy 981 MKQMS---YLMTLRGLNGYTLDTGTIOKISRDTGHVTOIKDSIKTTLTANYDLNHRIG 1037
Db 950 NKPISSGDYRYTMSGLIQRHKDSFAHDHYVSYDAEGLVTEGSSQYATFEYDNGRLIT 1009
Qy 1038 SOVTLATGMLTTTVEFDGLNREIGKLCDSGHTLIDIOQSWLKTQQLANRYKLVGL 1097
Db 1010 TTTKDTTSLSQLATKIEYDVFDRKIRSLISDFSIOV-ITLSYKNNQISQRLTSIDGV 1068
Qy 1098 ORTEOYSYDSRNLNOYKCGAECPTKXGHSIVTQNFPTDIYGNACHTTAPDGEDH 1157
Db 1069 MKKERIYDYNORISQYCEGEBQSPVDHTRVLSQYIYHDQWGNIRLONTIRADGET- 1127
Qy 1158 ATEKFNAPTPCQLEVHTHTPMDPNIRLKYPKAGRVINIT-DNHGNTENFTYDLGRL 1216
Db 1128 VDHFSQ-ADPTQIRI---TSDKOQIELSYDANG--NLTRBEKQOT--LIYDQNNRL 1177
Qy 1217 ---ONGQASV---YGYPLRLVSO--KTDTLQCELYRRETMVNEVRANGEMRL-- 1264
Db 1178 VQVYKDSKGNLVCOYQDALNKLTAQVLANGTIVNQ--YYASGVANVQLAGBETITWSSDK 1236
Qy 1265 -----RTGETTIAQGRASKVLLNGTDSQSYTILSDKONLSQBYASVYGGKHKSTAND 1316
Db 1237 QRLGHQSTKQGESYTYQ-----GTDHNSYTIASQNEHELMALSYTYGFRSLI--- 1285
Qy 1317 ASILGVNGERADPVSVTHLNGRYSDPTLMRPHTPDLSLPPGAGGINPYCYLADPIN 1376
Db 1286 SSILPGLNGAQVDVPTGVYFLNGRYRVFVPLMRHSPDSWSPFGGGINPYCYLADPIN 1345

```

```

Qy 1377 RSDPSGHLQSWQAMTGIQMGITAGLILTIAT--GGMATA-----ANGTAAATASTTALAF 1430
Db 1346 RIDNGLHLSAGGILGIVLAGIITIVGVSLGAPALISAGITIAAGALGALASTSAFAVTA 1405
Qy 1431 GALSVDTSITSYSGALIEDASPRASSIIGVNSMGMAAGLAEASAI-----KGGTK 1480
Db 1406 TVTGMAADSIGISALSEKDPKTAGILNISTGLVLSRGISALITTSLSIKSARBSGQ 1465
Qy 1481 LATH--LQAF-AEDGENALIKSTSESRIRKMYTRS 1513
Db 1466 VASTSVIGSVPIEFGEIA-----SRSSR-RMDIALS 1495

RESULT 3
Q4ZUT5_PSESY
ID Q4ZUT5_PSESY PRELIMINARY; PRT; 1578 AA.
AC Q4ZUT5;
DT 13-SEP-2005 (T-EMBLrel. 31, Created)
DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)
DE YD repeat.
GN ORFNames=Psytr 2044;
OS Pseudomonas syringae pv. syringae B728a.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=205918;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RG DOE Joint Genome Institute;
RA Chain P., Larimer P., DiBarolo G., Copeland A., Lykidis A., Trong S.,
RA Nolan M., Goldstein E., Thiel J., Malfacis S., Lapidus A., Deter J.C.,
RA Land M., Richardson P.M., Kyriades N.C., Ivanova N.;
RT "Comparison of two complete genome sequences of Pseudomonas syringae
RT pv. syringae B728a and pv. tomato DC3000."
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RA Loper J.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RA Feil H., Feil W.S., Lindow S.B.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000075; AAY37087.1; -; Genomic_DNA.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RHG_repeat; 6.
DR TIGRFAMs; TIGR01645; YD_repeat_2x; 2.
SQ SEQUENCE 1578 AA; 177054 MW; 51A4B576ABCC10C9 CRC64;

Query Match 18.2%; Score 1615; DB 2; Length 1578;
Best Local Similarity 30.4%; Pred. No. 1,le-78;
Matches 481; Conservative 251; Mismatches 680; Indels 168; Gaps 48;

Qy 14 MSDNNEFTQANNFTSAVSGVDPRGTGLYNIQITLGHVNGNLGPTLPILTSYSPINKT 73
Db 1 MTAISTVSHNAFNLSTVYGSGVDPRGTQYVAITLTPVKNTGLRGPMPLANNPILNRQ 60
Qy 74 DIGFGTGFNGSLSVYRKNSLSLSTGENYKVLET-DKTYKLOQKLDNIRFEDLKENC 132
Db 61 DSGFGGLWNLQLSQYDPGQIVSLSGEETRVGTSDQDMSEKIDSFHLKYE-DETH 119
Qy 133 YRIHKSQDELEVTFGNNAFDLKVPKGLNPAIGHAIYIDMNEATQPRIARIYDDLDGH 192
Db 120 FRVWHKSGVLVELLHLSGSGANRYALPVRYAPGHSYTLDYAFSGYQMLSEVYTD-SGQ 178
Qy 193 DIPLANLEYQGLIKTI-LTLPFGQKEGYRTFLPLNRQLNSHNFSLGNEPLTWSEGYT 251
Db 179 --LTLGISREDTAVLTIRMTTGEGPDATFKVVLGSGSHRVRI--ELPTDNAASWREYVS 233

```

```

QY 252 PIGKNGILGOWITSMAPGGLKETVNYSNNOGHFPOSANL PVL PVTYTLKQVPGAGQP 311
D 234 LIREH-----LCMTSVETPVGREGIEFTYDQ--GHQFAPARAPPLPVTNHTRTTREGQA 287
QY 312 AIOAEYS-----TSHNYVGGSGNGIW--NNKLDNLVGLMTE--YNYGSTESRYKDEGH 363
D 288 QMDVFSYQDSAGREBNFLAGLDIAMEDNGLDNLRYLAGELEYESTETLRLQVAGKP 347
QY 364 DQVRIERTNNHLLSECKOQNGYQTTETAAYALIGNPDSQBPQOLPKTKETMR 423
D 348 DAERSIRTRNOQFLLRRETRKVLTVQVTRYYIVPGQFPFTQPAVCQLPEVQTTMR 407
QY 424 -SADNSY-RSEITTFDESIGNPLTKYLKDKTKOKIISPTHMEYYPAGEVDCPEPY 481
D 408 LSSGSGVPRSEMVSTYDQKGNLLTQ-----QANGMEVVS---EWMYAAE--DCCPPDPE 459
QY 482 GFTFRVKKIQTPESEFKDPEKFIQYRSLISSQSHVTLKIEERHYSATQLNS--- 537
D 460 GFVFMKSSKIIIPASADYNAPVILTRRYTALPAVGGQLOQMLAPQSEITLLMQABEDE 519
QY 538 -----TLFQYNTDSEL---GRLLKQTECTGEMGKT-----YSVNHFTYTKODDTLQOS 585
D 520 YEBORLTYEYFDEBNMALLHGR-LKQRVIMGENAATSTEVASTPDESEYGR--TVLQTV 576
QY 586 HSITTHDNFTIHRQVRSRYTGRLEFSDTDKDIYQMSYDKLGRLLRRLNSGTP--YANT 644
D 577 QTVTFGPAKAKVTLLEBSLNGEBPLARDNNVIRYVYBSLRAVLSBTVAENDIYQAT 636
QY 645 LTYDEYELNLQDDNRPPFVITTTDVNGNQLNEFDGAGRHSQCLKSDGQ--KYYTI 701
D 637 RSYEQJCAKATDQATQ---TLIVKGVKCSHFQDLNVRVHERDADKPARVDKRRQT 693
QY 702 HTQOYDEGGRHHTSYDYLINGRQDIDPDVHLMSMSYMNQOINHTMSYVSEKIT 761
D 694 YAAVYDAMGLVNEFEDMWL---GOODLALKSVY-----EYDWMQBQRCVTGADNIKTPEE 746
QY 762 VDPF-TLATKQLOSNNSNVQTEKVTYTPS-----QQPQIQLTFEAGHLQSCHTLTLR 815
D 747 TDPFICTASQSPVQSWIEGTDGKGSFETETWNLPQPRSRKODAGTSVLSQYH 806
QY 816 DGMWRVKETDAI---GQCTIYQYDNVNRVIOITLPDGTIYNNKXAPSTDTLITDIRV- 871
D 807 DGLGRVREIIVGVNNGGRRTTYVDVDFRVEISTLPSFAIYRASAHSSEDLISIGIE 866
QY 872 -NG--ISLQCTPGLSRLTOSQSGRWATYAGANDQCSTYITTPGQGIHIOYQPEL 928
D 867 HNGKSIVLGEQSPGLDLVSSVYTGGRKRLVYDVGLIQ--FRYVTLPSKRIKIDYEKQV 925
QY 929 DD-----AVLQVANSNEITQOFSYNPVTGALLKAVA 958
D 926 GDEPMRVRQSEABEPDTRREVDPKQSGPAAAPARTAPBPVADYITDKQNALHSEE 985
QY 959 EGGSITPIYPSGRKIMENINDMK--MSYIMTLRGLENGYYTDLTGTIQKISNDTHGRV 1015
D 986 QGEILDRREYVSTGSLKSEKTSAGKTYDMFYSSQGLLSTYDVLQGEQVNRXYDTGRL 1045
QY 1016 TQIDSSIKTILAND---DLNRHIGSQVTLATLCHMLTTTVEPRGLAREIGRLCQSSG 1071
D 1046 BETRIGREYSTFADYVFGRLAKITTTQOTDSRREYTAISLELYDGLREVRKTF-DIDG 1104
QY 1072 HTLDIQSMLTKQOLANRIVKLANGVLQRTQEQYSYSRNRLNQKDCGAECTDKGHSIV 1131
D 1105 VEQGVQVYDDVDVQKTLISGTEILADENHNVTQSRLLTQYQCTGQRAPDPRGRLL 1164
QY 1132 TQNFYDIYGNITACHTTFADGTEHATFKFANPTDPCQLTEVNHHT--HPDMPDNIRLKYD 1190
D 1165 GQNFIFDQANMLTFYTRF--DQGSNNARY--FYEAGDPVQLSRVTNTNAAAYPPEINLKYD 1222
QY 1191 KAGRINITTDHNGENTFYDTLGRLOV-----GQSSYGYDPLANLVSQKDTTDLCEL-X 1245
D 1223 PDGNL--DIDAGRT--LKYDPLGRILEVGLTISAGIHHQYDPDQDLTE--TRGGGRDLRF 1277
QY 1246 YRETLVNEVRNGEMIRILRTGERTIIAQORASKVLLTGTDSQGSVILTSDKONLSQEAYS 1305

```

```

D 1278 YRDGELANQSGSGSTFMKGDYLLAEQGDRTLLATSASDSVGEVAADGVNRRWT 1337
QY 1306 AYGHKSTANDASITLGVNERADPVSGVTHLGVGYSYDPTLNRPFHTPDSLSPFGAGIN 1365
D 1338 AYGHASGDEPPHRGLDFNGELSEADTGQMLGVGAYSVLWRFNPSPDWSPPGBGGMN 1397
QY 1366 PYSYCGDPINRSDPSGHLQSWQAMTGIQMGIDAGLL-----TIATGMAIAAAGIA 1417
D 1398 AYAYVEGDVPMMDPFGH-----FGLFTPKLILYRLAKTTPITLTTSPGVKIPGL 1449
QY 1418 AALASTSTTALAFGLSVTSIDTISI-----VSGALEPASPKASSITL 1459
D 1450 TSGRKEVTLKSKITMDMDDLTDVAESYPARISAKIDNKPGIAATLTQKQTEAKDALN 1509
QY 1460 WVMGMCAGAL---AESATK 1476
D 1510 YLSAHVNOQGITAHARRAIK 1529

```

## RESULT 4

```

Q883V9_PSESM
ID Q883V9_PSESM PRELIMINARY; PRT; 1562 AA.
AC Q883V9;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE YD repeat protein.
GN OrderedLocustNames=PSPTO2239;
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STPAIN=DC3000;
RX MEDLINE=2834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolony J.F.,
RA Madupu R., Daugherty S.C., Brinkac L.M., Beaman M.J., Haft D.H.,
RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
RA Khoturi H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
RA Utecherack T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhouer S., Chatterjee A.K.,
RA Delaney T.P., Lazarewitz S.G., Martin G.B., Schneider D.J., Tang X.,
RA Bender C.L., White O., Fraser C.M., Collmer A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
RT Pseudomonas syringae pv. tomato DC3000."
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
DR EMBL; AB016853; AAO55755.1; -; Genomic_DNA.
DR TIGR; PSPTO2239; -.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; Rhs_repeat; 6.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 3.
KW Complete proteome.
SQ
SEQUENCE 1562 AA; 175713 MW; 8DC10DA1BFE37BF1 CRC64;

```

```

Query March 18.2%; Score 1614; DB 2; Length 1562;
Best Local Similarity 30.3%; Pred. No. 1.2e-76;
Matches 490; Conservative 239; Mismatches 700; Indels 190; Gaps 51;

QY 14 MSDNNEFTQANNTSAVSGGVDPRTGLVNIQITLGHIVGNGNLGPTLPLTLYSPINKT 73
D 1 MTSIVSHSNKFNMSVILQSGVDPRTOGYIVSINLPEVKSGLRGPVPLVLSNPLAVQ 60
QY 74 DIGGIGFNRGLSVYDRKNSILSTGBNRYVT-ETDKTVKLOQKLDNLRFKDKKNC 132
D 61 DSGGGLMNLQLSQYDGCGRIVISLGSGETFEVGSLSGDLQMLMPKXKDSFFRYKQ-DTR 119
QY 133 YRIHKGSDIEVLGFPNNMAFDLKVPPKLLNPAHAATYIDNFEATQPRLRIYDLDLGH 192
D 120 YRVVHKSQVHELEVLDSLGNRIALPVRIYSPBGHGLTHYASFGAYOMLSBVVD-DGQ 178

```

QY 193 DPLANLEYOGLIKITLTLFPQGEKGYRTBLRPLNRLNLSHNSFLGNEPLTWSFGTTP 252  
 Db 179 VILITTD-----SISVALLVGAAPKADAEPMLISGNRNVARLELPANKASRFTYYSI 234  
 QY 253 IKGKGLIGOWITMTABGKLKETVYNSNNNOGHHFPOSANLPLVYTLKOVYAGACOPA 312  
 Db 235 IKGHSC-----IASVDTFVGHEHEDVFYQDS--GHQFPLISAGREPLPRVTRHLTTGFLQPE 288  
 QY 313 IOAEYSI-----TSHNVVGGSGNGIW--ANKLDNIYGLM--TEVYGSTESRRYDKEGHD 364  
 Db 289 VDRVAYAKGAGREBNFLAGLIDJAMEBNGLDNIYRILGAPYLSSTETELRVNDV--- 345  
 QY 365 QIVAEKTYNNYHLITSECKOONGYIQTETAYAYAILGHNDSQSPQFOLPKTETWR- 423  
 Db 346 -VASIEVFPQFHLIAETTRQNLISLEVDRIYIEEGKPPDLPQNCQLPKEVRIITWRL 404  
 QY 424 SADNSY-RSEITETTFPESGNPLTKVTKDKTKQKIIISPTHEWEYYPAGEVNDCPPEYVG 482  
 Db 405 SPDSQVPRTEIVSDSDYSYGNLQTOANGVTETS-----EWSYSGE-DGCPDPDG 456  
 QY 483 PTPRVKIIOTPPYSEKDEKFTQYRY---SLISQSHVTLKIF---ERHYSATQ 533  
 Db 457 FVRLKAKSVVPAQSDGHALVLTARYRYKALPALASGQNLMLAASETLLOTTDGEK 516  
 QY 534 LNSTLFOYNTDKS-----BLGRLLKOTECTGKENGKT---YSVYHKEFTYTKODDTLQSH 586  
 Db 517 ELQOITTYIEDNPDYDAFOYGRIRHQSVTLEGLSTTDYRYDLQDDP---DQTVQTV 573  
 QY 587 SITTHDNFTIHRQVR---SRYTGRLEPSDITDIXDITVOMSYDLKRLITLTLN--SGTPYA 642  
 Db 574 QIVT--GFDMTQKVIKLEHSLFTGEPLLRNDNDVEIRYDNLRRAYSEIVSNKEBK 631  
 QY 643 NTLTYBELANLQDNRPPVITTTDVNGNQLRNEFGAGHNSQCLKSDSGDKFPTTH 702  
 Db 632 ATRHYEQLOAKVTDQAEORLF---DYKNVOTTSRFPGLGVITYEARADNDPVRRLD 688  
 QY 703 TOQ-----YDEQGHHTSTYSVDYLTNGRQOTDPKVLMSKSYDNMGQANTHMSYGS 757  
 Db 689 LRQTYEAYVDAWGDKVETSDYL-----DOOKALITNYPERIDDDQSLVTPGPGVT 741  
 QY 758 EKITVDITLTA-----KOLQSNNSNVQGEKVTY--TPSQOPIQITLPEAGHLOSCH 811  
 Db 742 TIEQDVEVTOASNGPIORRWTESNDGLQTSSEVETMLNLFDEPTRSVRLDRDLMLSEPV 801  
 QY 812 TLFR---DGDNRVAKETDAI---GQCTTYQVYDNARVYQITLPRGTVNRKYPASFTDL 865  
 Db 802 SLSEYQYDGLRGLVKEVSGPIRERSSTYGVDPDRVANTLPGAVRRRYAPHSGBDL 861  
 QY 866 ITDIRV--NGIS--LGOOTPDGLSRLOSQDGRVMAVYTSAGNDQCPSTVTPDQGFH 921  
 Db 862 PAMIGVHNGKASVLAGKEDGLDRITVSTTGGREBELSTSDLMQ--PKTVKLPSEGRQID 920  
 QY 922 YOYQPELDADVLQ-----VASNEITQOFSYNPYTGALLKAVAGESGLTPIYPSGRKLM 975  
 Db 921 YDVLPELGDPLKKTQSDVYARLKLTDYTYDPPGNARLISSEGBELQREHYSTGLKS 980  
 QY 976 ENINDM--KKMSYLTMLTGLNGEYTDLTGTQKISDTHGRTVQIROSSIKTTLATYDL 1032  
 Db 981 EORTSOGIENYMHYRYSRLPLSYIIVLQGOEQLSVYDDGRLAQTSILGEVSPFTYDTF 1040  
 QY 1033 NRHIGSOVTLATGMHTTVEPFGNLNREIGRKLCDSSGHTLIDQOSWMLTQOLANRYK 1092  
 Db 1041 GRPASITLDDSSNGGVVISLEVDAQGREAKRT--TINGANQOMVQYVDDVDQVAKETLS 1099  
 QY 1093 LNCVLQTEQYSDSRNRNLQYKCDACEPTDKYSHSVTQNTYDYGNIATCHTTPAD 1152  
 Db 1100 EGVAYIIEBHXYGLDQRLTQYDCSGKORPVPYGMVTSIQVDFDGNLMLTVTTTF-D 1158  
 QY 1153 GTEBHAFKXANPDPQOLTEVYHHTH-----PMPDNIRLKYDKAGVYN 1197  
 Db 1159 GGRRRAY--FYEGIDPAQLTRVNTQOLAMVNAARLIPVKNKDSITYPPIRLTYPDGML-- 1215  
 QY 1198 ITDNGHNTENTYDTLRL---QNGQGSV--YGYDPLNRLVSOQKTDLDCELYRETMVLN 1253

Db 1216 ITDEADL--LSTDPGLRLLEVSMRPAADVRRYRDPORLAGTGE-----QRFYRDTVLAS 1269  
 QY 1254 EVRNGEMIRLLRTGETTIIAOCRASKVLLTGTDSQOQVILTSDKONLSQEAYSAVGEKST 1313  
 Db 1270 QLGASQNSTYMRGEGYLLAEQCSDDLFTSISNSVLSVHPHDGVNSRSTVYVGHSSGD 1329  
 QY 1314 ANDASITGVNGERADPVSVGTHLNGGRSYDPTLMRPHITDLSLSPFAGINPYSYCLGD 1373  
 Db 1330 DPPAGRGVNGELHEHDTGMQLLNGGYRAVPLMRHSPDSMSPFEEGGILAAVAGEBD 1389  
 QY 1374 PINRDPESH-----LSWQATG-----IGMIGAILLTLTATG 1407  
 Db 1390 SVNGVDPPTGCMWRKRLRIFRTRLEKQAMRADKKEFRLIEDIKNEGLEGRAVQAARDL 1449  
 QY 1408 MAIAAAG--IAAIASTSTTALAFCALSVTSDITSVSGALEDAS--PKASIIIGWVS 1462  
 Db 1450 OAKSARKNKAYISISLVHGREKALE--AFITLKANTGDMVSSSREBSITPKAEASGSIS 1507  
 QY 1463 MGKGAAGLASAIKGTAKLATHGAPADEBNALIKSTSSSRKMGVTSLSREIYRN 1521  
 Db 1508 -----NLGFGSGSDRDVGLRVSSA---KSEVLARQD--FVRN 1540  
 RESULT 5  
 Q7NMU2\_CHRVO  
 ID Q7NMU2\_CHRVO PRELIMINARY, PRT, 1385 AA.  
 AC Q7NMU2;  
 DT 01-MAR-2004 (TREMREL. 26, Created)  
 DT 01-MAR-2004 (TREMREL. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMREL. 26, Last annotation update)  
 DE Nematicidal protein.  
 GN Name=Thsa; OrderedLocusNames=CV1887;  
 OS Chromobacterium violaceum.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 NC Neisseriaceae; Chromobacterium.  
 OX NCBI\_Taxid=536;  
 RN [1]  
 RP NCBIJCTDISE SEQUENCE.  
 RC STRAIN=ATCC 12472 / DSM 30191;  
 RX MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;  
 RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,  
 RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,  
 RA Alves-Gomes J.A., Andrade E.M., Aratipe J., de Araujo M.F.F.,  
 RA Acoulti-Filho S., Azevedo V., Baptista A.D., Bataus L.A.M.,  
 RA Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,  
 RA Bordignon J., Brito C.A., Brocchi M., Burtly H.A.,  
 RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,  
 RA Carvalho C.M.B., Cascado J.C.M., Cavada B.S., Chueire L.M.O.,  
 RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,  
 RA Fancinatti F., Farias I.P., Felipe M.S.S., Ferrati L.P., Ferro J.A.,  
 RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,  
 RA Gazzinelli R.T., Gomes B.A., Goncalves P.R., Grangeiro T.B.,  
 RA Gracitapaglia D., Gilsard E.C., Hanna E.S., Jardim S.N., Laurino J.,  
 RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,  
 RA Madeira H.M.F., Manfio G.P., Maranhao A.O., Martins W.S.,  
 RA di Mauro S.M.Z., de Medeiros S.R.B., Meisner R.V., Moreira M.A.M.,  
 RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,  
 RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,  
 RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,  
 RA Ramalho-Neto C.B., Reis A.M.M., Rigo L.U., Rondinelli E.,  
 RA Santos E.B.P., Santos F.R., Schneider M.P.C., Senarez H.N.,  
 RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,  
 RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,  
 RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Umenyi T.,  
 RA Vettore A., Waesem R., Zaha A., Simpson A.J.G.,  
 RT "The complete genome sequence of Chromobacterium violaceum reveals  
 RT remarkable and exploitable bacterial adaptability"  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).  
 DR EMBL; AB016916; AAG59561.2; -; genomic\_DNA.  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF05593; Rhs repeat; 6.  
 DR TIGRfam; TIGR01643; YD\_repeat\_2x; 3.





Best Local Similarity 44.6%; Pred. No. 8,66-76;  
Matches 342; Conservative 119; Mismatches 275; Indels 30; Gaps 17;

```

QY 741 YDMWGOJANTHMSGVSEKIVTDVITLTATKQLOSNNNVOTGEVITYTPSQOPIQITL 800
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 YDSWGQNHLLTVFSDGYGERSVYDPIITRRAT--LQPEGSQKLGQQLTEYNIAGLPKIVTQ 61
QY 801 FDEAGHLOSCHTLTRDGMVRKRETDALGQCTTIYQYDNVNRVIOITLPDGTIVNRKAPF 860
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 YDSGTEQGSANHYVDGLQARKETDELQVTLLEYDHFGRVOTLPENTIIQKSYAP 121
QY 861 STDTLITDIRVNGISLGOQTFDGLSRLTQSODGGRVMAVYTSAGNDQCPSTVITPDGQFI 920
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 STALITGISVNNFSMGNTQFDSLERLTETSGRTSAFSYENAS-SVPAAYATPAGETV 180
QY 921 HYQYQPELDVAVLQVANSNEITQOFSYNPVYTGALLKAV-AEGOSLPRIYPSGRKME-NI 978
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SYEYLKEIAGNAVKKISAPILKTWYDALTGAMTSATQAGMIRQMTYTPSGLLKNETSM 240
QY 979 ND--MKMSYLTMLRGLNGVYDITGTIOKISRDPHGRVTOIKDSSIKTTLANYDILNRH 1035
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 PDGAQKSTATVYSLAGAPOSTYDVGTVQRYDYDEHGRRIQIEDNDIKVSLGDAFGRF 300
QY 1036 IGSQVTLATGMLTTVEFPDGLNREIGRKLCDSSGHTLDIQGSMLKTQOLANRIVKLG 1095
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 TKQQATDKKTGAVLSTLTLYDILNREIKREISAGQSGLVIEQTYQRHLLKERITQGR 360
QY 1096 VLQRTQOYSYDSNRNLNOYKCDGACPTDKYGHISVTONFTYDIYGNITACHTTFADGTE 1155
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 TTLKEVFAYDSNRNLLEYTCNGEARPODPYGAIRHQTFSSYDALGNMTKTQTFDSG-R 419
QY 1156 DHATFEFANPTDPCQLTVEHHHTHPDMPDNIRLKYDKAGRVINITDNGHNTENTFYDTLGR 1215
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 NTAIYIY-SAIDPQOLKVNNDHSDYKPEITLEYDKAGMR--DEAGRT--LRYDALGR 474
QY 1216 LQ--NG--QGSVYGDPLNRLVSOKT-DTLDCELYYRETMVNEVR--NGEMIRLLRTG 1267
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 LQGVNAGAKGQGYAYDALNTLVSOVQVDEPIYDLYRADDLVGEARDOSSQTRYVXSN 534
QY 1268 ETTIAQ--QRAKV-LLTGDSQOSVILTSQKONLSQEA-YSAYGKHKSTANDASTILGY 1322
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 GCCCGQCTKQSSNTSRLTTTNOQSVLSVSEGNHAPQDCITYRGRTPQTEPSTVLG 594
QY 1323 NGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSCGDPINRSDPSG 1382
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 595 NGERLDPVSGTYHLNGYRAYNPILMRNCPDSWSPFAGGINPYACDGPINRVDPN 654
QY 1383 HLSMQAMTGMIGIAGLLTLTATGMAIAAAGIAAIAIYSTTALAAGALSVTSIDTIS 1442
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 HLSMQAELGIGIGVGLVLAFTAGTSIAAAGIASAIESASISLVVGTGLVADVASI 714
QY 1443 VSGALBDSAPKASSILGWVSMGMAAGLAE--SAIKGTGKLATHL 1485
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 715 ASGALBDANPQASATLGMISLIGGPGAVSGLATPARGKLLISGL 760

```

RESULT 8

```

ID 052880 COXBU PRELIMINARY; PRT; 774 AA.
AC 052880;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein orf 774.
GN Name=orf 774;
OS Coxiella burnetii.
OC Plasmid ORS.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Lautenschlaeger S., Jaeger C., Willems H., Baljer G.,

```

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y15898; CA475841.1; -, Genomic DNA.  
DR InterPro; IPR006530; YD.  
DR Pfam; PRO5593; RMS repeat; 6.  
DR TIGRFam; TIGR01645; YD\_repeat\_2x; 2.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 774 AA; 83774 MW; 918A543D7991BD8 CRC64;

Query Match 17.2%; Score 1530.5; DB 2; Length 774;  
Best Local Similarity 44.5%; Pred. No. 1,4e-74;  
Matches 341; Conservative 120; Mismatches 275; Indels 31; Gaps 18;

```

QY 741 YDMWGOJANTHMSGVSEKIVTDVITLTATKQLOSNNNVOTGEVITYTPSQOPIQITL 800
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 YDSWGQNHLLTVFSDGYGERSVYDPIITRRAT--LQPEGSQKLGQQLTEYNIAGLPKIVTQ 61
QY 801 FDEA-GHLOSCHTLTRDGMVRKRETDALGQCTTIYQYDNVNRVIOITLPDGTIVNRKAP 859
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 YDSGTEQGSANHYVDGLQARKETDELQVTLLEYDHFGRVOTLPENTIIQKSYAP 121
QY 860 FSTDITLDIRVNGISLGOQTFDGLSRLTQSODGGRVMAVYTSAGNDQCPSTVITPDGQFI 919
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 HSTASLITGISVNNFSMGNTQFDSLERLTETSGRTSAFSYENAS-SVPAAYATPAGET 180
QY 920 IHQYQPELDVAVLQVANSNEITQOFSYNPVYTGALLKAV-AEGOSLPRIYPSGRKME-N 977
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 VSEYLKEIAGNAVKKISAPILKTWYDALTGAMTSATQAGMIRQMTYTPSGLLKNETS 240
QY 978 IND--MKMSYLTMLRGLNGVYDITGTIOKISRDPHGRVTOIKDSSIKTTLANYDILNR 1034
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 PDGAQKSTATVYSLAGAPOSTYDVGTVQRYDYDEHGRRIQIEDNDIKVSLGDAFGR 300
QY 1035 HIGSQVTLATGMLTTVEFPDGLNREIGRKLCDSSGHTLDIQGSMLKTQOLANRIVKLN 1094
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 FTQQATDKKTGAVLSTLTLYDILNREIKREISAGQSGLVIEQTYQRHLLKERITQGR 360
QY 1095 GVLQRTQOYSYDSNRNLNOYKCDGACPTDKYGHISVTONFTYDIYGNITACHTTFADGT 1154
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 RTTLKEVFAYDSNRNLLEYTCNGEARPODPYGAIRHQTFSSYDALGNMTKTQTFDSG- 419
QY 1155 EDHATFEFANPTDPCQLTVEHHHTHPDMPDNIRLKYDKAGRVINITDNGHNTENTFYDTL 1214
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 RNTAIYIY-SAIDPQOLKVNNDHSDYKPEITLEYDKAGMR--DEAGRT--LRYDALG 474
QY 1215 RLQ--NG--QGSVYGDPLNRLVSOKT-DTLDCELYYRETMVNEVR--NGEMIRLLRT 1266
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 RLQGVNAGAKGQGYAYDALNTLVSOVQVDEPIYDLYRADDLVGEARDOSSQTRYVX 534
QY 1267 GETTIAQ--QRAKV-LLTGDSQOSVILTSQKONLSQEA-YSAYGKHKSTANDASTILG 1321
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 NGCCVGQCTKQSSNTSRLTTTNOQSVLSVSEGNHAPQDCITYRGRTPQTEPSTVLG 594
QY 1322 YNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSCGDPINRSDPS 1381
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 595 YNGERLDPVSGTYHLNGYRAYNPILMRNCPDSWSPFAGGINPYACDGPINRVDPN 654
QY 1382 GHLSMQAMTGMIGIAGLLTLTATGMAIAAAGIAAIAIYSTTALAAGALSVTSIDTIS 1441
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 GHLSQAELGIGIGVGLVLAFTAGTSIAAAGIASAIESASISLVVGTGLVADVAS 714
QY 1442 IVSGALBDSAPKASSILGWVSMGMAAGLAE--SAIKGTGKLATHL 1485
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 715 IASGALBDANPQASATLGMISLIGGPGAVSGLATPARGKLLISGL 761

```

RESULT 9

```

ID 042UT9_PSESY PRELIMINARY; PRT; 1682 AA.
AC 042UT9;
DT 13-SEP-2005 (Tremblrel. 31, Created)
DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)
DE YD repeat.

```



GN ORFNames=Psy 2040;  
 OS Pseudomonas syringae pv. syringae 8728a.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadales; Pseudomonas.  
 NCBI\_TaxID=205918;  
 RN NCBIOTIDE SEQUENCE.  
 RP NCBIOTIDE SEQUENCE.  
 RC STRAIN=8728a;  
 RG DOE Joint Genome Institute;  
 RA Chain P., Latimer F., DiBartolo G., Copeland A., Lykidis A., Trong S.,  
 RA Nolan M., Goldenberg E., Thiel J., Malfatti S., Lapidus A., Detter J.C.,  
 RA Land M., Richardson P.M., Kyriakides N.C., Ivanova N.,  
 RT "Comparison of two complete genome sequences of Pseudomonas syringae  
 RT pv. syringae 8728a and pv. tomato DC3000.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).  
 RN [2]  
 RP NCBIOTIDE SEQUENCE.  
 RC STRAIN=8728a;  
 RA Loper J.;  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP NCBIOTIDE SEQUENCE.  
 RC STRAIN=8728a;  
 RA Fell H., Fell W.S., Lindow S.E.;  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CP000075; AAY37083.1; -; Genomic DNA.  
 SO SEQUENCE 1682 AA; 186410 MW; CFPF95C541B81ECF CRC64;  
  
 Query Match 16.8%; Score 1490.5; DB 2; Length 1682;  
 Best Local Similarity 29.0%; Pred. No. 7.4e-72;  
 Matches 477; Conservative 237; Mismatches 715; Indels 217; Gaps 53;  
  
 QY 14 MSDNNEFTQANNNTSAVSGVDPRTGLNINQITLGHIVNGNIGPLPLTLSPANKT 73  
 DB 1 MAASTVSHSNALNMSCLKSGVDPRTGLNINISLPLQSDNLRGPGPRDLSSQLNTL 60  
 QY 74 DIGGIGFNGLSVYDRKNSLSTSTGENYKVIETDRT--VKLOQKLDMLRFEKDKEN 131  
 DB 61 DSGVGLGAMNQLSGYDPATQILSTSTETPRVDTSSNGQLTMREKLDITFHFVK-IDDE 119  
 QY 132 CYRIHKSGLIEVLTFGNNAFPLKVPKLLNPAGHAIYIDWN-PEATQPLNRIYDDL 190  
 DB 120 SYRAHSGGLVEILHLSGSKRMAMPVKLIAPGSHSIALKHTVFNSSYRLASITDGL 179  
 QY 191 GHQIPLNLEVOGLIKTI-----LTLFGQKGGVTELRFLNKLNS--IHNEELN 240  
 DB 180 -----QTLKTAASSTVVELDLHPDEGTGDTPLARFLMTLAGSDKRSRITLPT 228  
 QY 241 ENPLTMSFGYTPIGKNGILGOWITSMTPAPGLKETVYNNQGHFPOSANLPLVPLVT 300  
 DB 229 ENKASMPFEGLENGNOLC---VAHVETPAGSSENVY--QDEGHAFPSHGRMPVPRVT 283  
 QY 301 LMKOVPGAGPAIQAEYSY-----TSHNVYGGSGNGIW-NNKLDNLVGLMTEYVSGTES 354  
 DB 284 KHVLDPLNIAKIDVRYTYKGGQORSRNPGLAGLPIMEDNGLDNLKYVLQDYVYCTES 343  
 QY 355 RRYKDXKHQOIVRIETNNYHLJTECKQONVYIQTETAYYAIIGHNFDSPSPOL 414  
 DB 344 LWDVNRK---AVSIEKTEFRFHLQTLLEVTTQNNQOLVTTAYNIILOGEHYRSRQPNDCOL 399  
 QY 415 PKTETETWRSAD--NSYRSEITETTFPESGNPLTKVKDKKTKQKIISPTHWEXYPAGE 472  
 DB 400 PSEITTRMQLDAERTRTEIVDTYDSYGNLVHTPADG-----IEVSSW-YPAAGG 452  
 QY 473 VDNCPPEPVFTFVKKIQTTPYDSEKDDPEKFIQYRSLIGS-----OSHV 520  
 DB 453 -DCCPDADAEFVSLKEKVKVPASSQLGAPTLATRYQTLPALADSELPDIVPESRT 511  
 QY 521 TLKIEBHYSATQULNSTLFOY--NTDKSEL-GBLLQTECTYGENKTYSVNKKFYTK 577  
 DB 512 LLDQESD--GTCELOQLVLEYINQPDAPFLHGRGTGKNISLNDNTVT---YVESKSK 566  
 QY 578 --ODDTLQOQSHSITT--HDNFTIHSQVRSRYTGRLEPDTDKDITVQMSYDKLGRLLTR 633

DB 567 SRQLEVPVQIDITTTIGFDNAKISKQOQSILTGQVLL-TLEGVLEIRVYVDLNRLTKE 625  
 QY 634 TLNSGTPYANTLTYVDELNNLQDDNRPPIVITTTVDNGNOLRNEFDGAG-----HVS 686  
 DB 626 TIAPDSSDEASRELYLVLCSAACQAEQVIV---DARKVTRSVLDIGRAILIEERDHI- 681  
 QY 687 QCLKSDGDDKFTTIHQVDEGRHHTSVSYDLTNMGROTPDCKHLSMSYDNMGQ 746  
 DB 682 ----DSNDLOAMRIIHTAQYANMNNVOHERDYVMPAN--QQ---RATNTAYVDMNV 731  
 QY 747 IANTHMSYGVSEKITVDP-----TLTATKOLQSNVNVQKEXEYTYT---PSCQP 795  
 DB 732 QCVTISDHGQTHQYVDPIDNEHKGQVQKWTQVSSDAEPLIGRSSTWLMNGKPKIK 791  
 QY 796 IQITLPEAGHL-QSCHTLTRDGMDRVREKTDAGICTIYQYNNRVIOITLPDGTIVN 854  
 DB 792 IKIKTONADGEXQSGQTFLYDGLGCTEQTDSNMHTLPSYAMSMVYTRLPDGSSEAH 851  
 QY 855 RKYAPSTDLIDIRV---NGIS---LGGQTFDGLSRITQSDGGRVMAVYTSAGNDQC 908  
 DB 852 RQYAPHSNTELPGLKMMHSDGILTTLAGKQAFDGLGRLLTSQAGQRIEGETYEDGRQV 911  
 QY 909 PSTVITPDGQFIHYOYQPELDDAVLQVANSBITQOESYNPVTGALLKAVAEQSLTPYY 968  
 DB 912 -KTRKTAKGDDISYTNALATDQIVSSTAPDVTGFGVDVNSARLTSATNQCKRYVAYD 970  
 QY 969 PSGRLEKMINDMK---KMSYMTLRGLENGYTD-----LTGITQKI 1007  
 DB 971 AHQUTETKTKDLOGRTWKTVHRTSLQGRMYRTDVEQKDVNSVKGKVGKRIKGVKGV 1030  
 QY 1008 SR-----DTHGRVYQIDQSSIKITLNLVDLNRHIGSQVTDLATGMLTTPYFPGINREIG 1063  
 DB 1031 ETTIRYDDFERLENVQAGNEVITLLYDTLGQSEMTTCDDAAGRILNRRKKTIDQGGEL 1090  
 QY 1064 RKLDSGHTLIDIQOSWLTQOLANRIVKLVGLVQRTQYSYDSRNRLNOKYCDGACPT 1123  
 DB 1091 RTQIAGNHPRITLQOQMDTGMLQMSRHLQEAQNTLLETFKYDARGSLTVIYSGSLPV 1150  
 QY 1124 DKYGHSLVTONFTYDIYGNITACHTTFADGTEDHATFKFANPTDP-----COLTVHH 1176  
 DB 1151 HASGRALIKQTFIFDSIDNMNIQTTEEPADASTEMAFPHYKEDDPDPLSRDRCOLLGITY 1210  
 QY 1177 THPDMPONIRLKYDKKGRVININQNTENFTYDILGRL---QNGG---SVYGYPLN 1230  
 DB 1211 RPRGTPDPPTFRYDANGN--QLDDEHGN--RLYYDSQSLRLRYEKPAGBPISTAYVGH 1266  
 QY 1231 RLVSOKTDLIDCEL--YYRETMVNEVRNGEMTRILRTGETIIAQO---RASKVILTGT 1285  
 DB 1267 HLAFTTNGS--DSRIARFYQOQQLSSTVQDORRQFLYIDQPVGQQTIGSPATILLTD 1325  
 QY 1286 SQGSVILTSQKUNLSQEAAYSAYGKHSSTANDASILGNGERADPVSGVTHLNGYRSYD 1345  
 DB 1326 ANOSVLAFFQDDKLRARVAYSAYGERHSDDALLVAGANGSEICEDKTMWYLLGNGYRAYNP 1385  
 QY 1346 TLMREHTPDSLSPGAGGINPVSYCLGDPINRSDPSH----- 1383  
 DB 1386 GMRPFSPDLSLSPGAGGVAPYTCLENPAMRDPGHDASSQSGRLRPDEDAIPAE 1445  
 QY 1384 --LSWQAMTIGIGI---AGLLLTATGMAIAAGGIAAIAASTSTTA----- 1427  
 DB 1446 GGMGIQIWMILANVIGVITNAGAVATVASF--LAAQSVVTVLGMNTMTQTAAYVATGL 1503  
 QY 1428 LAFGALSVTSDITSYISGALF--DASPKASISIGWWSMGK--AAGLAESAIKGKTALATH 1484  
 DB 1504 LAAQTVLGAATAATYTAAYGAVKGETAFRLGEYLAVALPIDIAGVFRVSAIKAAAS- 1562  
 QY 1485 LGAFAEDGENALLKSTSE--SSRIKMG 1509  
 DB 1563 -----KASSRVSSETASRVSVG 1579

RESULT 10

Q88LP7\_PSEBK  
ID Q88LP7\_PSEBK PRELIMINARY; PRT; 1632 AA.  
AC Q88LP7;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocustNames=PP1882;  
OS Pseudomonas putida (strain KT2440).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=160488;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22423060; PubMed=12534463;  
RX DOI=10.1046/j.1462-2920.2002.00366.x;  
RA Nelson K.B., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,  
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
RA Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,  
RA Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,  
RA Hance I., Chris Lee P., Holtzapple E.K., Scanlan D., Tran K.,  
RA Morazzaz A., Uettermann T.R., Rizzo M., Lee K., Kosack D., Moestl D.,  
RA Weiler H., Luder J., Scjepandic D., Honeisel J., Straetz M., Heim S.,  
RA Kewitz C., Bisen J.A., Timmis K.N., Duesterhoeft A., Thiemler B.,  
RA Fraser C.M.;  
RT "Complete genome sequence and comparative analysis of the  
RT metabolically versatile Pseudomonas putida KT2440.";  
RL Environ Microbiol. 4:799-808(2002).  
DR EMBL; AE016780; AAN67501.1 -; Genomic\_DNA.  
DR TIGR; PP1882;  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RBS repeat; 7.  
DR TIGRfam; TIGR01643; YD repeat 2x; 4.  
KW Complete proteome: Hypothetical protein.  
SQ SEQUENCE 1632 AA; 182290 MW; F07C2A5820160401 CRC64;  
Query Match 15.9%; Score 1416; DB 2; Length 1632;  
Best Local Similarity 28.4%; Pred. No. 8e-68;  
Matches 472; Conservative 232; Mismatches 734; Indels 226; Gaps 52;  
QY 14 MSDNNEFTQANNPTSAVSGVDPRGTGLNLTQTLGHVGNNGNIGPTPLPLTSLSPINKT 73  
DB 1 MSTSSLSHNSALNMFMSFMSGVDSPRTQYTLSTIKLPGLGANYFSGPELISLNTPLMLMI 60  
QY 74 DIGIGIFGNFGLSVYDRKNSLSLSTGENYKVI-ETDXTVKLOQKDKDNLRFKDKL--- 129  
DB 61 DSGMGKNSLILTPAPHTQYITTSGBSPKVTGSSGSRLOMQEGRLLHFFHYEPAPRG 120  
QY 130 -ENCYRIHKSGLIEVLTFGNNAFDLKVPKKLNPAGHAITYIDWN-PEATQPLNRIYD 187  
DB 121 GNARYVVRHSGSLVEILEMGSANGRIMALPEVITYAATGRIDLQYQPNSSYMLLSISD 180  
QY 188 -----DUDGHIPLNLEYOGLITLILTFPGQEGYRTFLRNQLNS--HNF 236  
DB 181 EREBILIEIEDSDSRIELRERYQG-----DNGQPVALYANN-LITDWTSTI 227  
QY 237 SLGNENPLTWSFGYPIGNGILGQIMTAPGALKETVYVYNNQGHFPOSANLPLV 296  
DB 228 VLPFHELASWRLVAV--NGLL--CVSKVERPTARRELYY--QDRHLRFGDAR-PPL 280  
QY 297 FYVTLMKVPGAGQPAIDAEYSYTS-HNYVGGSGNGIV-NKLDNLVY-GLMTEYNGSTE 353  
DB 281 PRVTHVIEPRGGQAFQRTYTYPGILNPLFGAGIGMSDNGLDMLYESKRYDEYQYVE 340  
QY 354 SRRYDKGHDQIVALEETVYNNYHLLTSECKQNGYIQTTERAVYAIIGHNDSQSPQFO 413  
DB 341 TLR--DEGG-TALRDIITFTFNRFLITSTFYVNNCAHEV-TWQYNIDVPPNQQVSTLQ 396  
QY 414 LPTKTEITWRSADNGYSR--EITETTFDESGLPTKVIKDKTKQIKLISPTHEWEYYPAG 471  
DB 357 MPRIQRTKSLADNITRSLRLETVEITTYDSSGIIRKLANGVTER-----QEWYTGQA 449  
QY 472 EVDNCPPEPYGFTFRVKKIIGTPYDSEFKDPEKFIQYRYSILG--SQSHVTLK----- 523

DB 450 E-DVYGDANGFVRHKSKTITTPASSGRGQAPLTOHYRYKALAPLAGNAITLNPIVER 508  
QY 524 ----IEBRHSATQNLNSTLFQYNTDKSELGRLLKQTECTGNGKTSYV---HKFTYTK 577  
DB 509 SETTLTAHPAHPLKEEKIYIYLDAPKSSL-----RGRRYQEVVKKKLETTT 556  
QY 578 Q-----DDTLQOSSHSTTHDNFTIHSQVRSRYTGLFSPDPTKDIYVQNSYD 625  
DB 557 QYQFNSLIDPLGGHQVLETKTLFGYDGAQRSTYQRRSLHGEKLYELNENGVTQWAD 616  
QY 626 KLGRLTLRTLNSTGTPPANTLTYYELANLQDDNRPPVITTYDNGQLANEPDGAARHV 685  
DB 617 ALRRVTEERSVSPETPEAKKRYDYQLCASDADLARARV--TYNARGITTELELDGLRPT 673  
QY 686 SOCLKID--SDGDKFYTIHTQYDEQGRHHTSTYSVLTNGRQOTPDCKYHLSMSKSYDN 743  
DB 674 RESRDVLEARKPPAGFYITLAIQIDAVGNRIQESVTWLO--AQD---YHLVTKRYDD 727  
QY 744 WQGIANTHMSYGVSEKITVDPI-----TLTATKQLQS-NSNNVQ----- 781  
DB 728 WGEQCCITIGPDVGEQHNVLDPFGNADHQAI RYSWREGRLKSPQYQNRNGMNVRSRQ 787  
QY 782 ----TGKEVTTYPSQOPIQITLPDEAGHLQSCHTLTRDQWDRKRETDALGCTIYOYD 837  
DB 788 ARAISGKTEITWMLFNKPVLRKRLDALGELIGRRYSYDLAGRTLTETDERMHTTAFSYD 847  
QY 838 NYNRVIGITLPDGTIVNRKYPASTPDLIT-----DIRVNGISLQOTPDGLSRLTQS 890  
DB 848 AWGRMLTTEQPNKTLIRTYAHSABDLPTRLVYTPANVQLPARQIAQAFDGLDRILIGT 907  
QY 891 QDGGRWAVYYSAGNDQCPSTVITPDGQFIHYQYQPELDAVLQVANSNETTOQFSYNPVT 950  
DB 908 TYGBRTEFLFRDG-ESIPIQRINPAGETIEQYINLQLTPEISNTAPEEHSAPYDPS 966  
QY 951 GALLKAVAEQSLPTIYPSGRKLKMNINMKMSYMTLRGLENGTTDLTGITQKLSRD 1010  
DB 967 ARLISDNRQCARRFENKANKOQLAEHMEW-KRDGKWSRHSSTLQDLRLNTEHPYGED 1025  
QY 1011 T-----HGRVTOIKDSSIKTLYNYDDLNRIHIGSQVTDLTAIGHMLTTVEERDGLNREIGR 1064  
DB 1026 TTHERYDAQSLVSTLQQLQAEFKYDDLGRIELITSHDRASAOALFKIERYDDQDRYKR 1085  
QY 1065 KLCDSGHTLIDIQSMKLTQOLANRIVKLVGLQRTQYSDYSNNRLNOKYCGAECPTD 1124  
DB 1086 TWKPGQPERQETVMDKDLISRTLQVGVSVSLVEKFGVSHARLNMYCTSPDQPRD 1145  
QY 1125 KYGHSYVTONFTYIYGNITPACTTTPADG--TEBDAETKFNPTPCQLTEVNHHTPRMD 1183  
DB 1146 ALGRSIAMQVFNFAVNNILTVTSFTGGPAERATFTIME-RDPCQLRLEYRPPRTAP 1204  
QY 1184 NIRLKYDKAGVINITNHNGBTENFTYDLGLRL-----QNGQSYVGYDPLNRLVYSQKTD 1239  
DB 1205 NPESYDANG--NLTRDE-QARPIRYDSQRRLLGLANDSGAPDYDAGGLVSRPAG 1260  
QY 1236 KQNSQEARYSAYGKHKSTANDASITLGYNGEADVDVSGVTHLGNQYRSVDPPLTMEFHPDS 1355  
DB 1321 AGSTRAVRYTAVGERHADDPILGTLGYNGEALDDPSGMYLLGSGYRAVNPVLMRPHSDPA 1380  
QY 1356 LSPFGAGGINPYSYCLDPIRNSDPSGHL-----WQ-----AWT 1390  
DB 1381 LSPFGAGGALNYGCGQGNPITFRDPTGHSYIGYSGQSRSLADLMSYSTWRKALGALQWT 1440  
QY 1391 GIGMGI-----AGLLTLATGMA-----IAAGGIA-----AALASTS--- 1424  
DB 1441 GIGIGILFAAVVASVAVVVTGVAAPALIAAAMAAGGIIISGAGALSAVGAVIASVSLAV 1500  
QY 1425 ----TTALAPALSTVSDITSIVSGALEDASPKASSILGWSMGGAAGLAEASAIKSGTJK 1480



```

Db 1501 GIKLAAVATATLSVTG--TALQTEALISGHEKKNIGITILNYSALIGLAVGAMQILAK 1558
Qy 1481 ---LATHGAPAD--GEMNLKSTSSSRKMKVTSLDREIV 1519
Db 1559 IPNLWAGSYTTTITDPLINSKTLSELK-----MAKQII 1596

RESULT 11
045948 COXBU
ID 045948 COXBU PRELIMINARY; PRT; 709 AA.
AC 045948;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE 045948.
OS Coxiella burnetii.
OC Bacteri; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
NCBI_TaxID=777;
[1]
RN NUCLEOTIDE SEQUENCE.
RA Thiele D., Willems H., Haas M., Kraus H.;
RT "Analysis of the entire nucleotide sequence of the cryptic plasmid
RT QPH1 from Coxiella burnetii."
RL Eur. J. Epidemiol. 10:413-420(1995).
DR EMBL; X75356; CAA53129.1; -; Genomic_DNA.
DR PIR; S38241; S38241.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RNS_repeat; 8.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 4.
KM Plasmid.
SQ SEQUENCE 709 AA; 79416 MW; A37789A2A72CCF86 CRC64;

Query Match 13.7%; Score 1214; DB 2; Length 709;
Best local similarity 40.8%; Pred. No. 2.2e-57;
Matches 287; Conservative 101; Mismatches 280; Indels 36; Gaps 16;

Qy 554 KQTECTGKNGKTYSVNKHKTYYTKODPTLQOSHITTHDNFTIRSGVRSRYGRFSDT 613
Db 3 RRAVLTSSKGGKTYQONQTTFAISSQAMHLQKIDFTGDDGKITSIHQGRYSGRLLST 62
Qy 614 DTDIDIVTQMSYDKLGRLLTTRTLN--SGTPYANTLTYDYELNNLQDDNRPVITTTDNGN 672
Db 63 DELGANTYQYDEIGRLTQTVANASTTYASTTYSISLELDARGKTKATKTTTDDKGN 122
Qy 673 QLRNEFDGAGRHSQCCKSDG----DGKRYTHITQYDEGCHHTSTYSYDYLTVNGROQ 727
Db 123 QLRITYDGLGRNLKQERLDDAAVSGKTGWTYTHQOYDALGRSKTITQDVL-----R 177
Qy 728 TDDEKHA---LMSK--SYDNWQILANTMSYVSEKITVDPITLTATKQLQNSNNVQ 781
Db 178 LDBEVGAGSVLSTSKMHWDSWQNHLYFSDDYQSRSYDPIITRAT--LQPESSQK 235
Qy 782 TGSKVTYTPSOQPIQITLFDGAGHLQSCHTLFRDGDVRYKEDALQGCCTIYOYDNYR 841
Db 236 LGGQLEYNLAGRPIKTYQYDSQTEGSAHYEDIGQLRKETDELQTLLEYDHFGR 295
Qy 842 VIQITLPDGTIVNRKVAFFSTDLITDIRVNGISLGGQTFDGLSRKLTQSDGGRVAYTY 901
Db 296 VTQTTLPENTIIQGSYVAFHSTASLTGISVNNFSGMGTFSLERLTETTSGGRTAFSY 355
Qy 902 SAANDQPSVITPDGFIHQYQPELDVAVQVASEITQPSYNNVTGALKAV--AEG 960
Db 356 ENNS--SVPAVAVTPTGTSVSEYVKEIGVAKISABEIIQTDYDALTGAMTSATQAG 414
Qy 961 QSLTPIYPSGRKME--NIND--MKMSYLTMLTGLNGYTDLTGIIQKISRTHGRVT 1016
Db 415 MINGMTYPSGLKNERSMFPGAAQKSTATTYSLAGRPOSTYDVFYGTQYVYDDEHKKI 474
Qy 1017 QIYDSIKITLNTDNLNRHIGSVQVTLATGHMLTTVEFDGLNREIGRKLCDSSGHTLDT 1076
Db 475 GIEBNDIKVSLDYDAFGRPTKQATDKCTGAVLSTLTLYDNLNREIGRISASQSVLVI 534

```

```

Qy 1077 QGSLKTKQOLANRIVKLVGLQRTQYXSDSRRLNLYKCDGACPTDKYGHSTVQNF 1136
Db 535 EGYQRNHLKERTQGRITTLREKMPAYDSRNRLIETQNGEARPDPPGKAIHQTF 594
Qy 1137 YDIYGNITACHTTTPADTETHATFKFANPTDPCQTEVHHHTPPMPNIRLYKDKARVI 1196
Db 595 YDALGNITKQTPDSGQ--RNTATYIY--SALDPQLKVNNDHSPPEKITLEYDKARMI 652
Qy 1197 NITDNHNTENFTYDYLGRLO--NG---QGSVGYDPLNRLVSO 1235
Db 653 R--DEAGRT--LRYDALGRLOVNGAGKGGQVAYDVYLVTLVSO 692

RESULT 12
0881P2_PSEPK
ID 0881P2_PSEPK PRELIMINARY; PRT; 1290 AA.
AC 0881P2;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=PP1887;
OS Pseudomonas putida (strain KT2440).
OC Bacteri; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=160488;
[1]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22423060; PubMed=12534463;
RX DOI=10.1046/j.1462-2920.2002.00366.x;
RA Nelson K.E., Weinel C., Paulsen I.T.,
RA Martins dos Santos V.A.P., Foute D.B., Gill S.R., Pop M., Holmes M.,
RA Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,
RA Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,
RA Hance I., Chris Lee P., Holtzapfel B.K., Scanlan D., Tran K.,
RA Moazzaz A., Ultebreck T.R., Rizzo M., Lee K., Koback D., West D.,
RA Medler C., Lauber J., Stjepandic D., Hohelsel J., Straetz M., Helm S.,
RA Klewitz C., Biesen J.A., Timmis K.N., Duesterhoeft A., Thiemler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL; AE016781; AAN67506.1; -; Genomic_DNA.
DR TIGR; PP1887; -.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RNS_repeat; 5.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 4.
KM Complete proteome; Hypothetical protein.
SQ SEQUENCE 1290 AA; 146578 MW; 7081A3F10B287F92 CRC64;

Query Match 11.7%; Score 1037; DB 2; Length 1290;
Best local similarity 27.7%; Pred. No. 2.5e-47;
Matches 375; Conservative 196; Mismatches 578; Indels 204; Gaps 52;

Qy 14 MSDNNEFTQANNFTAVSGVDPRTGLNINQITLGHIVNGMLGPTLPLTSSPLNKT 73
Db 2 MSTSVSHSNAFNFSSTIESGVDPRTGYVTSIRLPELQNDLQGFELALFSPSPLNGE 61
Qy 74 DIGFGIFNGSLGVYRKNSLISLSTGENYKVI--ETDKYKLOQKLDNLRPEKDLREN- 131
Db 62 DSGFGKGMN--LQLTQYRKIVTLSSGETYKIKTGKSVYORLMEKGRKQRF--DLYEDP 117
Qy 132 -----CYRIHKSGDIEVLTFGNNAFPLDKVPKLLNPAAGNAYIDW--NFEATQRLN 183
Db 118 PGCGAARFVEYHRSGLVEVLFWGSGEGEVALPVELHSLPGRHILHSYLPFGCHRLS 177
Qy 184 RIYDDLGDHIDILNLEYGQILKTLTLFRGQGEVGRTERF--LNRQNSIHNFSLGN 240
Db 178 EVDQSD---VLRQSRSDNSRVELLCPYSGGDDGAPLARVAMTLSESNDRVSEIILPT 233
Qy 241 ENPLTWSFGYTPIGKNGIIGQ--WITSMTAPGGIKETVYNNNNQGHNFPOSANILPV-LPY 298

```

Db 234 ANQARWFTY-----EDVLGVLCVRECTPPYGVVERVY---QDAGHKFPSSAARDKXLP 286  
QY 299 VTLMKVPGAGQPAIQAEISYT-SHNTYVGGGSGNINMKN-LDNIYGLMTEYNTGSTESRR 356  
Db 287 VTRHEIDPRGQAKVVRVEYPTGTHNFIIGGSSISMSDDGIDNIYKRPEDYTYKSID1-- 344  
QY 357 YKKEGHDOIVR-IERTYNNYHLLTSECKOONGYIQETTERAYVALIGHNPDOSPQOLP 415  
Db 345 ---OEVARQSVRTTRTFNRPMLTEQATACGDKLQALPRTYADNAG-NESQEPYQLP 400  
QY 416 KTKETWRSDN--SYRSEITETTFDESQNPITKVIKDKTKQKISBSTMERYPPAGEV 473  
Db 401 HDETFQMSLISNPERQREKRVTRYSHGVLTRLLPNQVLETNV-----WYSABEEG 453  
QY 474 DNECPREYGTFRPK-KIQ-----TPYDSEFKDDPEFTQRY 511  
Db 454 DE-----HGFVRNLTCTVQPAATGHAATAITGAFYRALTPPGS-----YLQOPW 500  
QY 512 SLIGSGSHVTLKIBERYSAATQLN--STLFQ-VNTQKSELGRLLKQTECTKG-ENGKTY 567  
Db 501 RLUESR-----ISESSGAPETFEKISKLYQESASERFSYGRVQQTVSIPGVGGSPF 555  
QY 568 SVNKEFTYTKQDD--TLQOSHSTITHDNFTIHSQVRSRYTGRLLPSDTKCDIYQMSYD 625  
Db 556 DTLTHYGSLLPDBERALTQVEKLVGVDOESKTIYLRHALDTGEBPLMLDNGVETRYYD 615  
QY 626 KLGHLTRTNSGTPYANTLTLYDELNNLDQDNRPPIYTTDVNGQLRNEPFGAGRIV 685  
Db 616 ALRVTREVVAPGRPEABRNRYKLCAYDNEQAQMAVYDKQVETHTL--LDGLSRPI 672  
QY 686 SQCLKSDSG--DGEFTYIHTQYDEQGRHNTSYSDVLTNGRQOTDPDKVHLSMSKSYD 742  
Db 673 FEEBADDSATYAAALRPYAKRDELQDLVEBTEIDMLGG-----LLELTSHSYD 725  
QY 743 NMGG--IANTHWSYGVSEKITVDPITLT-----ATKQLOSNNNVQTEKEVTTYPSQ 793  
Db 726 DMGQRYAVLNPDDGTQVEE---IDQVASTDGPVRAHMEVY---HSRSIGITETWMLFE 779  
QY 794 QPIQITLFDKAGHQ-SCHTLTRDGMRVKRETAIG---QCTIYQVDMNVRVQITLPD 849  
Db 780 KPVRIERFALDSTISILQVNDYDGLGLSREAGSGAGGVVREYRDAFRAVEERLAD 839  
QY 850 G-TIVNRKYPAPSTDTLITDIRV-----GISLQCFPEGRSLTOSODGQWVAUTYSA 903  
Db 840 RVNTVYRTYAGHSNDLPVSIKQNTBSAAMVLGEQVDSGLERPTVALTGERBEQFEPR 899  
QY 904 GNDQCPSTVITPDGQFIHYQYQPEL-DDAVLQVANSNEITQGFSTNPVTGALLKAVAGQS 962  
Db 900 GERQ-PHVVKAPDGRTRIEYQYRPALGEBPVLRLVSGK-EAKYVDLKNARLTHCEBPDD 957  
QY 963 LTPY-----YSGRLKMN-----INDMKMSYIMTLRGLNGCYTDLTGTIQKISPDTH 1012  
Db 958 ENSGTYLDRSHPLNSGEVRESRTYVGEAFSMTYDYSFRSLRAVVDLGGQTYDPFDV 1017  
QY 1013 GRVTOIKDS-----IKTTLNYDNLNRHIGSQVTDLATGQMLTTT 1052  
Db 1018 GRLEKTIHAPEPKTRRYRLQARARQLLLESTFGYRQGRMASITTTDASTGHALATL 1077  
QY 1053 VEPDGLNREIGRKLCDSSGHTLTDIQOSWLKTQQLANRIYK-----LNGVLTORTQYS 1104  
Db 1078 LEYBEFDEIRLTF-DFGDTYQTLAQDYDEPDLKSRILKERPKGSDSOATLLRHEYYQ 1136  
QY 1105 YDSNRLNQYKCDAGCEPTDKYGSIVYONFTYDIYGNITACHTFAGCTEDHATFKAN 1164  
Db 1137 YDRGRRLQIYTCDEAPVDDSGQITARQIFGFDLNDILSVITYRPGSQORTLYEKEN 1196  
QY 1165 PTDPCQ-----LTEVHHTHPDMPD-----NIRLKXKAGRYINITDNHGN 1204  
Db 1197 -SDPAQMSRIIPDALLLETYVDVHDLBELFKI VAKDPQLADLHDNGGNL--ISDBGR 1253  
QY 1205 TENFTYDTLGR-----ONGGSGYGYDPLNRL 1232  
Db 1254 V--LTYDGLNRLRLVETPDGERCRYNVDPENIL 1284

RESULT 13  
Q883W6\_PSESMS  
ID Q883W6\_PSESMS PRELIMINARY; PRT; 982 AA.  
AC Q883W6;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE YD repeat protein.  
GN OrderedLocustNames=PSPT02231;  
OS Pseudomonas syringae (pv. tomato).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=323;  
RN [1]  
RP NOCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=DC3000;  
RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;  
RA Buell C.R., Joarard V., Lindeberg M., Selengut J., Paulsen I.T.,  
RA Gwyn M.L., Dodson R.J., Debey R.T., Durkin A.S., Kolonay J.F.,  
RA Madupu R., Daugherty S.C., Zafar N., Zhan L., Liu J., Yan Q.,  
RA Nelson W.C., Davidson T.M., Zafar N., Zhan L., Liu J., Yan Q.,  
RA Khoui H.M., Fedorova N.B., Tran B., Russell D., Betsy K.J.,  
RA Utechtack T.R., Van Aken S.B., Feldblyum T.V., D'Ascenzo M.,  
RA Dang W.-L., Ramos A.R., Alfano J.R., Carlinhour S., Chatterjee A.K.,  
RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,  
RA Bender C.L., White O., Fraser C.M., Collier A.;  
RT "The complete genome sequence of the Arabidopsis and tomato pathogen  
Pseudomonas syringae pv. tomato DC3000."  
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).  
DR EMBL: AB016853; AAC5747.1; -; Genomic\_DNA.  
DR TIGR: PSP02231; -;  
DR InterPro: IPR006530; YD.  
DR Pfam: PF05593; RNS\_repeat; 6.  
DR TIGRFAMs: TIGR01643; YD\_repeat\_2x; 2.  
DR Complete proteome.  
SQ SEQUENCE 982 AA; 106800 MW; 00E7CA6B93E19B78 CRC64;  
  
Query Match 10.5%; Score 932.5; DB 2; Length 982;  
Best Local Similarity 29.7%; Pred. No. 7.9e-42;  
Matches 285; Conservative 142; Mismatches 357; Indels 175; Gaps 33;  
  
QY 642 ANTLTYDEELNNLDQDNRPPIYTTVDVNGNQLRNEFDGAGRVSQCLKSDGGEFTYI 701  
Db 31 SSTIGRYD-----DMNGR--CCTTTDDN-VQYREYSDPIGSDVHK-----GPIQKT 74  
QY 702 HTQCYDEQGRHNTSYSDVLTNGRQOTDPDKVHLSMSKYDNMGQILANTHWSYVSEKIT 761  
Db 75 WKQSGDPEGR-----ISGRSEF-----MLNT-----FGKPRIR 103  
QY 762 VDPITLTATQLOSNNNVQTEKEVTTYPSQPIQITLFDKAGHLOSCHTLTRDGMRV 821  
Db 104 ----TLTAGTGRSHRTWSRSRLTY--TEQEL-----SRQTFYDGLGRC 144  
QY 822 RKEFDAIGCCTIYQYDNYNVIQITLPDGTIVNRKYPASTDTLITDIRV---NGIS--- 875  
Db 145 TEGPDALQOSTFSEYDWSMWSSTLADGSVIRNSYAPQSSSELATLLEVNHQGTRTV 204  
QY 876 LGQTFDGLSRLTQSGQGRVWATYSAGNDQCPSTVITPDGQFIHYQYQPELDAVLQV 935  
Db 205 AGTQKFDGLRVRVQTKYGDREVEQFNVDAGEMQ--PRSHYTAGLDINIFTYATLTDQIFSS 263  
QY 936 ASNEITQGFSTNPVTGALLKAVAGSGSLTPIYPSGRLKXNENIMDKMSYLV-----T 989  
Db 264 TAPDETAKFYDKTSABRLIATNPQGRTRYRYVDVHNLTGSTWNL--LQAMETRHQSS 321  
QY 990 LRGLNGYVTDL-----TGTIQKISRDTGRVTOIKDSIKTTLNVDNLNRHIGSQVTDIA 1044  
Db 322 ILGRPIKRTDLKKEAAGAEFRVYDYLGRIRFINGSNLRKRTIDYDVLGQCYKXVATELQ 381  
QY 1045 TGHMLTTVEFDGLNREIGRKLCDSSGHTLTDIQOSWLKTQQLANRIYKLVNGVLTORTQYS 1104

```

Db 382 AGTGVIIIDMEYDDOGGEIILRTQTASNOALTLTQTMAVDGLKTRLDQAGSPILHETFS 441
Qy 1105 YDSRRLNOKCKGAEPTKCHSITQNFYDIYGNITACHTTPADGEGHATFEKFN 1164
Db 442 YDERGRLLTVNYLSSLPDELQREMTQIFSPDELNITLCQRFDFGTGSEBAFCKGS 501
Qy 1165 P-----TDPQOLTEVHHHTHMDPMDNIRLKDQGRVINTINDNGNTENFTYDILGRL--- 1216
Db 502 PGDDKHKDRQQLSIATVPKRPDPFPSTIDANGN--QLKDEHGN--SLHYDSQSLQLQV 557
Qy 1217 -QNGQG--SVYGYDPLNRLVYSOKTDLDCBL--YYRETMVNEVRNGEMIRLLRTGETI 1271
Db 558 AETGAPISQYRYDGHQNLVATR-DGNESEILRFEHQSLSTVQEDQRTQYHLHGEQPL 616
Qy 1272 AQC---RASKVLLTGTDSSQSVILTSKQNLSCQAYSAYGKHKSTANDASITLYNGRAD 1328
Db 617 GQGIIVDAEQTLLLTIDANQSVMEFQOQLRKAVYSAYGERHSEBALSTAGNGEVR 676
Qy 1329 PVSGVTHLNGYRSYDPTLRFPDLSPPGAGIIPYSYCLGDPINRSDPSGH----- 1383
Db 677 AANGWILGNGYRAYNPLRFRHSPDLSPFAGGVNPTTCLGNPLALMDPTGHDAAG 736
Qy 1384 -----LSWQ-----AWTIGMGIA-----GLLTITAGMAIAAGGI----- 1416
Db 737 TGLRRLPDEGALPMQGGGDIMGVGVIGVFTVGLVAATITATLGATPVTGPTVLGI 796
Qy 1417 -----AAALASTTTALAG-----ALSVSDITSYSGALEDSPPASSILGVN----- 1462
Db 797 SMTASAAAVSTVSTGLIYGTALTAATTANTVAIYNN-----DQTAGEGVGLGIAAV 851
Qy 1463 -----MGGAAGLAEAIKGGTKLA-----THLGAFAEDGENALLKSTSESSR 1505
Db 852 PVGLVFGAGAVARAAVAAAARAAANAGTIGVRSVRIG-LAAAGARRTISMAASAR 909

```

## RESULT 14

Q87VG6\_PSESM

ID Q87VG6\_PSESM PRELIMINARY; PRT; 1669 AA.

```

AC 087VG6;
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE YD repeat protein.
GN OrderedlocusNames=BSPT04970;
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=323;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=DC3000; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RX MEDLINE=22834015;
RA Buell C.R., Joardar V., Lindberg M., Selengut J., Paulsen I.T.,
RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R., Daugherty S.C., Brinkac L.M., Beaman M.J., Haft D.H.,
RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
RA Khouri H.M., Petrova N.B., Tran B., Russell D., Berry K.J.,
RA Utecherack T.R., Van Aken S.E., Feldblum T.V., D'Ascenzo M.,
RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
RA Bender C.L., White O., Fraser C.M., Collier A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
Pseudomonas syringae pv. tomato DC3000."
Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
DR EMBL; AEO16853; AAO58398.1; -, Genomic_DNA.
DR TIGR; PSEPT04970; -.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RNS_repeat; 5.
DR TIGRfam; TIGR01643; YD_repeat_2x; 2.
KW Complete proteome.
SQ
SEQUENCE 1669 AA; 184268 MW; BC668B6C1E3FB7EE CRC64;

```

Query Match 10.4%; Score 920.5; DB 2; Length 1669;

Best Local Similarity 25.9%; Pred. No. 8,1e-41; Matches 440; Conservative 212; Mismatches 662; Indels 387; Gaps 84;

```

Qy 152 AFDLVPKQLNDPAGHAIYIDNPEATOPRLNRIYDDLGDHIDPLNLEYGQ---IKTI 208
Db 72 AFGQLLPSEVTSFQGGTLSTIEMRNOQGFLLKKI---SSDNLV---LSTYSAPSSTSTV 127
Qy 209 -LTLFPQGGEGYRTELRLANROLNSHNSLGNENPLTWSEFGTTPGKNIILQWITSMT 267
Db 128 EISVWPTDAAVYRLLELSYALTLGLRIEKNLKVQKLYGC---ADPTLRVLRRIE 184
Qy 268 APGLKETVYNSNNQGHHPQSANLPVLPYTLMKQVPGAGOPAIQAEYSYSHNYVG 327
Db 185 EBDGSIELVYTRRG--GMPPYRQ--PLPCVTLHSLFPAACSNITDHFYGTNYLGF 241
Qy 328 GSNGINNNKLDNIYGLMTEYNYSSTSRKYDGH-----DOVRIERTNNYHL 378
Db 242 SEPP--DAHQNRLY-----YERLELRLELVDEGYVQILRONPHIIVSTRHAFNNHL 292
Qy 379 LTESCKQONQYIQTETAYAYAILGNHFDSPQ--FQLPK--TKETWRSADSYSEIT 434
Db 293 QVRE-----DLQVQFAEKVISMEPANASPGKVFGLPYKTIYDYDLSHPNTERITV 346
Qy 435 EITFDESGNPLTKYIKDKTKTKIISPTHEYYYPAG-----EYD-----NC 476
Db 347 QTLAYNNIGQLTKSI-----AVDGVTEMLYTPDQGGGLDLSIAEKPLSKDVLTLTC 400
Qy 477 PPEPYGTFREVKKIIGTPYDESEKDP-----EKPIQYRY-----SLIGS 516
Db 401 PKVSEGY-----MPVKEVYHDPAYPKSGQITAYAYQARENPNQSVLVPSTVV 452
Qy 517 QSHVTLKIEBRHSAQOLNSTLPQYNTDKSELRLLKQRECKG-----ENGKTY--S 568
Db 453 LTGVTL-----DTTMLPSLM-----EGRNALIEQKIVISRSLPDVAVENTTAKES 499
Qy 569 VVHKFTYTKQDDTLQOSHSTITHDNFTIHR-----SQVSRYYRFLPSDTDK 616
Db 500 VYQNSMLGMQRNLT--TSMLYDNDNSVGVTVAAEAGKIISSRIFSRLSGRPLSER-RD 556
Qy 617 DIVQMSYDKLGRLLR--TLNSGTPYANTL--TYDEI--NNLQDNRPPVIYTTDVN 670
Db 557 GLFRFYVHDSIGRLVQERGTBAGMKADAVETTVYSIAEGIQ-----VTVEAE 608
Qy 671 GNLRLNEFDGAGHVSQCLKSD--GDKFTYTHQOYDEGRHHTSYSYLNGRQTD 729
Db 609 -QQVRLYDGLQRPVWVAIKRTLLPDAFCVIRIEYDGDATNQTLY--DYLGGGLRRTK 666
Qy 730 PDKVHLMSKSYDNMQIANTHWSYG--VSEKITVDPITLTATKROLQSNNNYQTKREV 787
Db 667 DAREPAAVDASKLAW--MADYTRDAGILINEQYIGADSGAQLRQLSGRLDLSHTALLE 725
Qy 788 TYTPS-----QQPIQITLFEAGHLQSCHTLTRDGMQVRKRETDALGOCTTYQYDYN 840
Db 726 TLAPSTABASTDRTIERT--FDERRLIKIRT-----SMTSEHC--IEYDELE 770
Qy 841 RVQITLPPQSTVYNRKKAAPTSTLLTDIRVNGISLGGQTFDGLSRLTSGQDGRWAYT 900
Db 771 RAAVALIAPDRTIRKRYHOLS--DYITQLNVGSTVIGTQMTAAAGQTVGE---LTYE 824
Qy 901 YSAGNQCPESTVTPPGQFIHYQYQPELDPAVQVANSNETQO-----SSYNPVT- 950
Db 825 FPGGS---ASTVRPDKTILLESASVADGHTATISINKKHTQVIGSPNVLYTVDPVSV 881
Qy 951 -----GALLKAAVAGGSLPIYY--PSGRKLMENINDMKKMSYLLWTLRGLNGYTDLTGT 1003
Db 882 PSAEAMSLTSSPQSIGLTSTIGTSPRGSQAQEMTRSLKG-----RLLTN--TAVDGR 932
Qy 1004 IOKISPDHGRVYQIQQSSIKTTLANDDLNRHGS-----QVTLANGHMLTTVPERPGIN 1059
Db 933 QMKVFYDYLDVVRV---ILGELHYHLYMSAFGBLPQRTVVAQASERLDVRFYMDAFG 988
Qy 1060 REIGRKLCSSGHTLDIQGSLWLTQOLANRIVKLVGLQRTQEQSYDSRRLNQYKC--D 1117

```

